

Ulaganja u sport i korelacija s uspjehom: Analiza i vizualizacija sportskih podataka

Peršić, Matea

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:195:631697>

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

Download date / Datum preuzimanja: **2024-07-11**



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

Matea Peršić

Ulaganja u sport i korelacija s
uspjehom: Analiza i vizualizacija
sportskih podataka

Završni rad

Mentor: doc.dr. sc. Lucia Načinović Prskalo

Rijeka, rujan 2023.



Rijeka, 4.4.2023.

Zadatak za završni rad

Pristupnik: Matea Peršić

Naziv završnog rada: Ulaganja u sport i korelacija s uspjehom: Analiza i vizualizacija sportskih podataka

Naziv završnog rada na engleskom jeziku: Investments in sport and correlation with success: Analysis and visualization of sports data

Sadržaj zadatka: Zadatak završnog rada je izvršiti analizu sportskih rezultata pojedinih država, odnosno razmotriti korelaciju uspjeha pojedine države u vidu osvojenih medalja na Olimpijskim igrama s količinom ulaganja novčanih sredstava te države u sport po stanovniku. Uz pomoć alata za pripremu, obradu, analizu i vizualizaciju podataka analizirat će se i vizualizirati odabrani skupovi podataka, a skupovi podataka i korišteni alati i tehnologije bit će detaljno opisani.

Mentor

Doc. dr. sc. Lucia Načinović Prskalo

Voditelj za završne radove

Doc. dr. sc. Miran Pobar

Zadatak preuzet: 21.4.2023.

(potpis pristupnika)



Sažetak

U današnje vrijeme digitalnog doba, kroz različite izvore, kao što su društvene mreže, senzori i internet transakcije, generiraju se ogromne količine podataka. Podatkovna znanost (*engl. Data science*) je interdisciplinarna znanost koja se bavi prikupljanjem, obradom, analizom, interpretacijom i prezentacijom podataka s ciljem donošenja informiranih odluka. Najčešća mjera koja se koristi u području podatkovne znanosti je korelacija (*engl. Correlation*), a ona nam govori o stupnju povezanosti neke dvije odabrane varijable. Primjena izračuna korelacije u podatkovnoj znanosti može biti u više različitih područja kao npr. analiza tržišta, financijska analiza ili sportska analiza rezultata. U ovom radu izvršit će se analiza sportskih rezultata, odnosno razmotriti korelacija uspjeha pojedine države s ulaganjem u sport u vidu osvojenih medalja na Olimpijskim igrama s količinom ulaganja novčanih sredstava te države u sport po stanovniku. Također će se opisati korišteni skupovi podataka te će se prikazati statističke metode koje su se koristile pri analizi odabranih skupova podataka. Osim same analize, u fokusu će biti i alati i metode za vizualizaciju podataka. Na kraju će se iz dobivenih rezultata i vizualnih formata izvesti zaključci i rezultati analize.

Ključne riječi: analiza podataka, vizualizacija podataka, korelacija, sport, sportski podaci, olimpijske medalje

Summary

In today's digital age, huge amounts of data are generated through various sources such as social networks, sensors, and Internet transactions. Data science is an interdisciplinary science that deals with the collection, processing, analysis, interpretation, and presentation of data to make informed decisions. The most commonly used measure in the field of data science is correlation, which provides information about the degree of association between two selected variables. The application of correlation calculations in data science can be done in many different fields, such as market analysis, financial analysis, or sports score analysis. In this paper, the analysis of sports outcomes is performed, i.e., the correlation of an individual country's success with investment in sports in terms of medals won at the Olympic Games with the amount of money per capita invested in sports by that country is considered. It also describes the datasets used and presents the statistical methods used in the analysis of the selected datasets. In addition to the analysis itself, tools and methods for data visualization are also presented. Finally, the conclusions and results of the analysis are drawn from the obtained results and visual representations.

Key words: data analysis, data visualization, correlation, sports, sports data, olympic medals

Tablica sadržaja

Sažetak	3
Summary	4
1. Uvod	6
2. Opis korištenih alata i tehnologija	9
2.1. Visual Studio Code	9
2.1. Tableau	10
3. Opis i obrada skupa podataka	13
3.1. Učitavanje u Pandas DataFrame	16
3.2. Korištene statističke metode	18
3.3. Ulaganja u sport i korelacija	24
Ulaganja u sport	24
Korelacija	24
3.4. Metode vizualizacije	28
Vizualizacije u Tableau	28
4. Zaključak	39
5. Popis slika	40
6. Popis tablica	41
7. Literatura	41

1. Uvod

S obzirom na svakodnevni rast količine podataka (*engl. Big data*), može se reći kako su analiza i vizualizacija podataka postali ključni alati za dobivanje najvažnijih informacija iz tih podataka. Unatoč sveprisutnosti velikih količina podataka u istraživanjima i analizama, ne postoji „dogovor“ koja se količina može nazvati velikim podacima (*engl. Big data*). Internet stranice prate svaki korisnikov korak, pametni satovi prate otkucaje srca, broj koraka koje napravimo kroz dan, a pametni automobili vrlo pomno prate vozačeve navike (Grus, 2019). U tim masivnim skupovima podataka leže odgovori na bezbroj pitanja koja do sada nisu bila postavljena. Iako ima i već poznatih pitanja, s njima se razvija pozitivan pogled na velike podatke i poslovnu analitiku jer analitika zalazi „iza površine“ velikih podataka kako bi se olakšalo donošenje poslovnih odluka temeljenih na predviđanju (Baesens, Bapna, R. Marsden, Vanthienen, & Zhao, 2016). Ipak, nisu nam sve informacije korisne, pa trebamo znati razumjeti i izvući najvažnije informacije iz podataka koje smo obradili. U tome nam pomaže analiza i vizualizacija podataka koja se sastoji od dva ključna koraka: analize podataka, koja predstavlja proces primjene statističkih i računalnih tehnika za prikupljanje, obradu, interpretaciju i prezentaciju podataka, te vizualizacije podataka koja predstavlja proces prikazivanja dobivenih rezultata u obliku grafikona, karti, dijagrama ili drugih vizualnih formata, koja se koristi kako bi se korisnicima olakšalo razumijevanje informacija dobivenih pomoću analize podataka. Opće je poznato kako ulaganja u sport uvelike utječu na same sportske rezultate i na važnost donošenja sportskih odluka. Takve odluke imaju vrlo važan utjecaj na razne elemente modernog sporta, kako na pojedince sportaše, tako i na timove, klubove ili reprezentacije. Može se pretpostaviti kako zemlje koje više novaca i sredstava ulažu u svoje sportaše, imaju bolje rezultate na natjecanjima, kako na lokalnoj razini, tako i na regionalnoj tj. svjetskoj razini.

Također, ulaganja u sport jedna su od ključnih faktora za razvoj sportskih talenata i poboljšanje sportske infrastrukture. Sport kakvog danas poznajemo otvorio je mnoge puteve za poslovno ulaganje nekolicine tvrtki, kao i velik broj multinacionalnih kompanija. Jedan od najrazvijenijih načina za poslovno ulaganje naziva se sponzorstvo ili sponzorski odnos (Kos Kavran, Kralj, & Ratković, 2020.). Na primjer, Zagrebačkom klubu „GNK Dinamo Zagreb“ sponzori su velike kompanije poznate diljem svijeta: „Coca-Cola“, „Adidas“ i „Wolt“. Stabilna financijska podrška omogućuje igračima, trenerima i upravi pojedinog kluba mogućnost napretka, rad s boljom opremom i stručnijim trenerima te daje veće mogućnosti nagrade sportaša. To se vidi na primjeru nogometnog kluba „GNK Dinamo Zagreb“. Ovaj nogometni klub jedan je od najpoznatijih klubova u Hrvatskom nogometu kroz više od 100 godina, najtrofejniji je nogometni klub u Hrvatskoj te je poznat i na razini svjetskog nogometa u „Ligi Prvaka“ (*engl. UEFA Champions League*). Cilj sponzorstva je zadovoljiti obje strane tj. davatelja sponzorstva (sponzor) i primatelja sponzorstva (sponzorirani) (Kos Kavran, Kralj, & Ratković, 2020.). Sponzoru je u fokusu iskoristiti talente i potencijale primatelja sponzorstva za ispunjavanje svoje marketinške politike, a cilj primatelja sponzorstva da uz financijsku podršku može lagodno i učinkovito obavljati svoje sportske obaveze. Možemo zaključiti kako klub ne bi imao takve rezultate bez vjernih i kvalitetnih sponzora.

Za ulaganja u sport vrlo su bitni i odnosi s javnošću. Izgradnja imidža svakog individualnog sportaša ili kluba ovisi o sportskim uspjesima, ali i o ponašanju i odnosima koje izgradi s javnošću (Kos Kavran, Kralj, & Ratković, 2020.). Velik utjecaj na to imaju i društvene mreže tj. svugdje prisutno *online* okruženje, gdje je snaga medijskog prostora jača nego ikada. U takvim slučajevima dolazi i do izgradnje odnosa s navijačkim skupinama koji su jedan od bitnijih elementa uspjeha sportaša (Kos Kavran, Kralj, & Ratković, 2020., str. 94). Postojanje samih navijača pozitivno utječe na igrače te garantira sigurnu i pouzdanu bazu podrške. Bitno je da se odnosi s navijačima temelje na zdravim odnosima unutar sportske organizacije, a posebno s igračima pojedinog kluba zbog mogućnosti privlačenja većeg broja sponzora i financijske pomoći. Klub ili sportska organizacija gradi dobre odnose s navijačima na nekoliko načina od kojih su neki popusti na članarine, besplatne članarine, davanjem navijačkih rekvizita navijačima i popusti na navijačke rekvizite (Kos Kavran, Kralj, & Ratković, 2020., str. 94). Pozitivnim imidžem

individualnog sportaša ili sportskog kluba, stječe se brojnija publika te rezultira atraktivnijim sportskim događajima.

Cilj ovog završnog rada usmjeren je utvrditi postoji li korelacija između ulaganja u sportove i sportaše na postizanje sportskih rezultata, bilo da se radi o svjetskim, međunarodnim ili državnim natjecanjima. Prateći taj cilj, analizira se kako financijska podrška i investicije u sportske programe te općenito sportske sustave utječu na sportske rezultate i postizanje uspjeha. Početna pretpostavka odnosno hipoteza je da, što su veća ulaganja u sportaše i uvjete u kojima sportaši borave ili treniraju, to su bolji i njihovi rezultati i uspjesi. Ulaganja u sportove dakako ne moraju imati utjecaj samo na fizičko zdravlje sportaša pojedinca, već i na društvene i ekonomske učinke, kao što su pomoć pri širenju svijesti o zdravim navikama te promoviranje određene zemlje kroz sport.

Analiza i vizualizacija podataka ima nekoliko ključnih primjena od kojih su neke:

1. Predviđanje potresa
 - Koristeći analizu podataka, moguće je predviđati potrese putem prikupljanja podataka iz seizmičkih senzora ukopanih u zemlju. Nakon prikupljanja podataka, oni se obrađuju i analiziraju kako bi se izgradio prediktivni model sposoban predvidjeti buduće potrese.
2. Prikaz klimatskih promjena
 - Analiza i vizualizacija podataka igra ključnu ulogu u prikazu klimatskih promjena i njihovih posljedica na okoliš i društvo. Podaci o temperaturi, padalinama i drugim klimatskim faktorima mogu se prikazati na različite načine kako bi se ilustrirale promjene tijekom vremena i njihov utjecaj na našu okolinu. Vizualizacija omogućuje preglednost i intuitivno razumijevanje složenih obrazaca i trendova klimatskih promjena.
3. Personalizirana medicina
 - Analiza podataka sve više pronalazi primjenu u medicini, posebno u području personalizirane dijagnostike i liječenja. Podaci o pacijentu, uključujući genetski kod, medicinske zapise i druge relevantne čimbenike, mogu se pažljivo analizirati kako bi se stvorio individualizirani plan liječenja. Kombinirajući tehnike analize podataka poput strojnog učenja i rudarenja podataka s bogatim medicinskim informacijama, mogu se otkriti skriveni uzorci, identificirati rizici i predvidjeti odgovor na određene terapije.
4. Analiza sportskih podataka
 - Pomoću analize i vizualizacije podataka mogu se predvidjeti budući sportski rezultati, analizirati prethodni rezultati individualnog sportaša ili kluba te donositi odluke o budućim natjecanjima klubova npr. hoće li ulagati više ili manje novaca u pojedinog sportaša, hoće li klubovi zadržati iste igrače ili dovesti nove te isplati li im se to s obzirom na rezultate koje su postigli prethodnih sezona.

Ukratko, analiza i vizualizacija podataka igraju ključnu ulogu u medicini, seizmologiji, klimatologiji, sportskoj analizi te mnogim drugim područjima. Kroz različite alate u analizi i vizualizaciji podataka, moguće je otkriti skrivene uzorke, identificirati trendove i prezentirati složene informacije na pristupačan način, potičući tako informirane odluke i inovacije u različitim područjima.

U idućem poglavlju biti će opisani korišteni alati i tehnologije koje su se koristile za izradu zadataka za ovaj završni rad. U poglavlju broj 3. opisao se i obradio skup podataka koji je korišten za obradu i vizualizaciju podataka. Također, opisane su različite korištene statističke metode i objašnjena je korelacija. Napravljene su različite vizualizacije na temelju statističkih izračuna i izračuna korelacije.

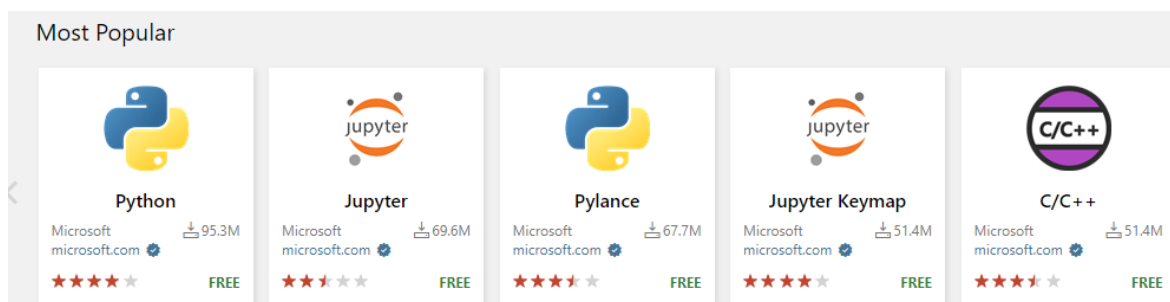
Kada bi se sve objedinilo, može se zaključiti kako će se u ovom završnom radu nalaziti brojevi i vizualni prikazi onoga što se želi prikazati, a to je jesu li sportski rezultati i ulaganja u sport međusobno povezani.

2. Opis korištenih alata i tehnologija

Softverski alati i tehnologije koje su korištene za izradu zadataka za završni rad su *Visual Studio Code* i programski jezik *Python* za analizu podataka, te softverski alat *Tableau* sa svojim intuitivnim sučeljem za prikaz i vizualizaciju dobivenih rezultata zbog lakšeg razumijevanja i interpretacije istih.

2.1. Visual Studio Code

Visual Studio Code¹ jednostavan je, ali moćan alat izdan 2015. godine pod MIT licencom². Služi za uređivanje izvornog koda i dostupan je za različite operacijske sustave: Windows, macOS i Linux (Visual Studio Code, 2023). Također ima ugrađenu podršku za JavaScript, TypeScript i Node.js i nudi mogućnost korištenja različitih programskih jezika kao što su C, C++, C#, Java, Python i mnogi drugi. Za pisanje izvornog koda u tim jezicima, potrebno je instalirati proširenja koje VS Code nudi. Na slici 1 prikazana su neka proširenja za VS Code.



Slika 1 - Najpoznatija proširenja u VS Code – u, snimka zaslona napravljena sa: (Microsoft, Extensions, 2023)

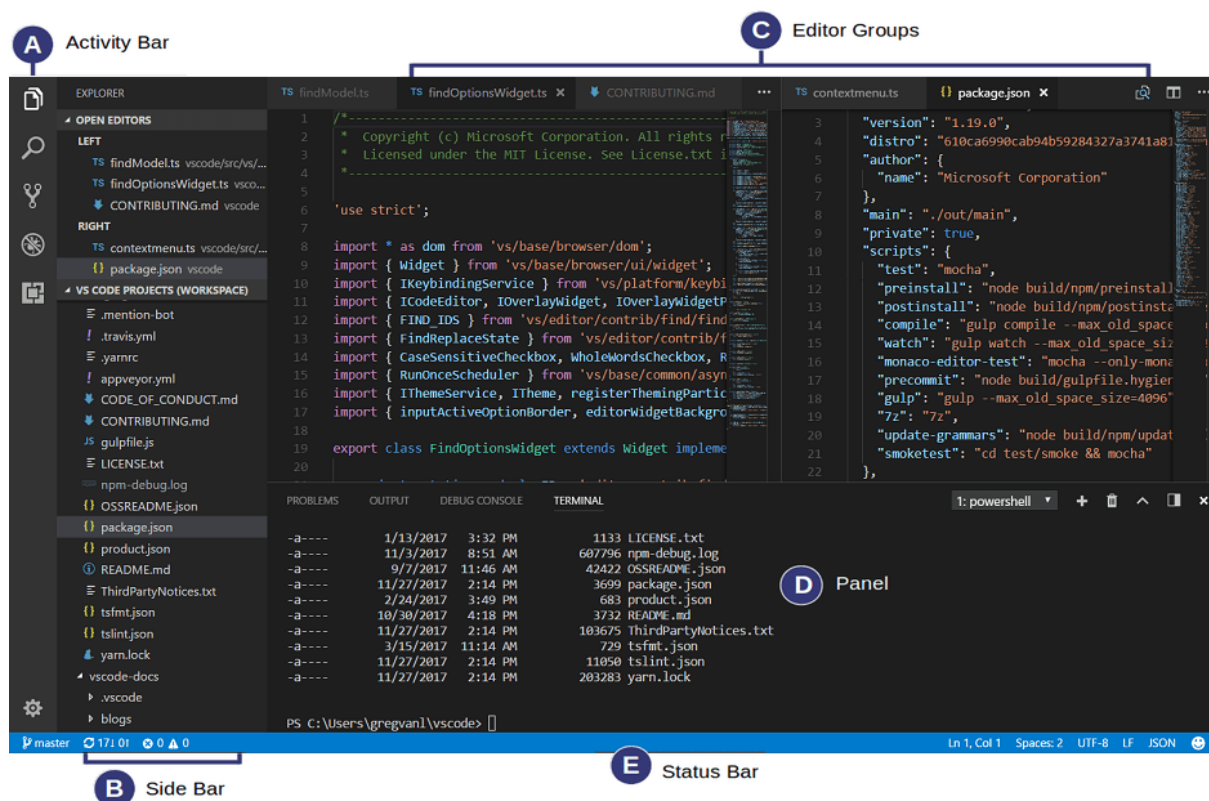
Njegovo je sučelje intuitivno i lagano za korištenje te je podijeljeno na 5 glavnih dijelova:

- uređivač (*engl. Editor*) – ovo je glavni dio za uređivanje datoteka, više uređivača se može otvoriti jedan pored drugog, vodoravno i okomito,
- glavna bočna traka (*engl. Primary Side Bar*) - sadrži različite elemente, kao na primjer 'Explorer' koji pomaže tijekom rada na vlastitom projektu,
- statusna traka (*engl. Status Bar*) – ovdje se nalaze informacije o trenutno otvorenom projektu i datotekama,
- traka aktivnosti (*engl. Activity Bar*) – smještena na krajnjoj lijevoj strani, pruža informacije poput broja promjena na Git-u,
- dodatni radni prostor (*engl. Panel*) – dodatni prostor koji se nalazi desno od središnje trake, sadrži različite kartice kao što su 'Problems', 'Output', 'Debug Console' i 'Terminal'.

Slika 2 prikazuje prethodno opisane dijelove sučelja Visual Studio Code – a.

¹ Često se naziva i VS Code

² Daje se dopuštenje svakoj osobi koja dobije kopiju softvera i pridružene dokumentacije da radi sa Softverom besplatno i bez ograničenja, uključujući prava na korištenje, kopiranje, modificiranje, spajanje, objavljivanje, distribuiranje, podlicenciranje i/ili prodavanje kopije Softvera (The MIT License, n.d.)



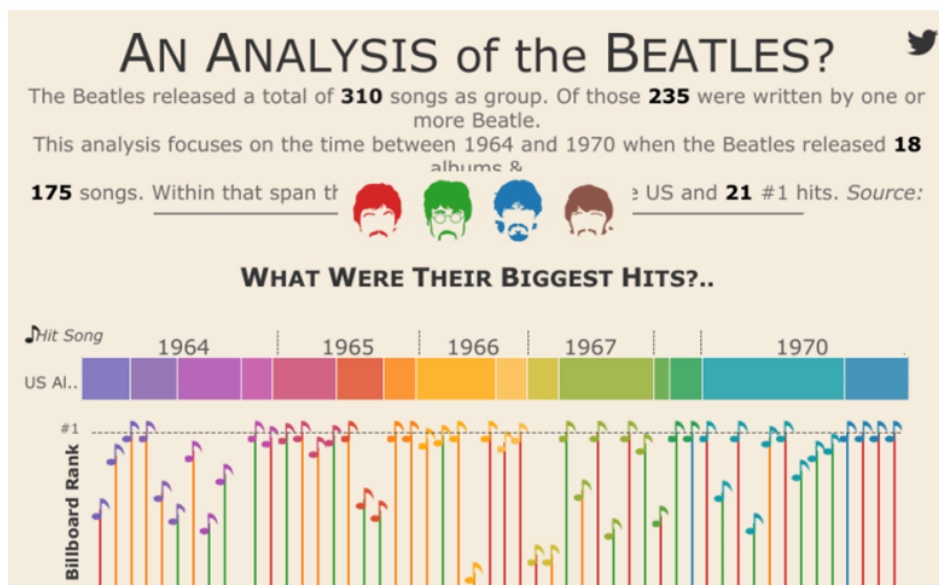
Slika 2 - Korisničko sučelje VS Code-a, preuzeto sa: (Microsoft, User Interface, 2023)

U istraživanju koje je provodio Stack Overflow 2023. godine, VS Code proglašen je najpoznatijim razvojnim alatom za programere zbog svoje jednostavnosti i intuitivnog sučelja. Od skoro 87 tisuća ispitanika, pokazalo se da ga 73.71% ispitanika koristi za učenje programiranja/profesionalno programiranje (Visual Studio Code, 2023).

2.1. Tableau

Tableau Software je američka tvrtka koja je sa svojim radom započela 2003. godine i od tada je njihova popularnost neprekidno raste (Tableau Software, 2023). Tableau je alat za vizualizaciju podataka koji se koristi za analizu podataka i izvješćivanje iz velikih količina podataka (Simplilearn, 2023). Često se koristi za analiziranje kod poslovne analitike (*engl. Business Intelligence – BI*)³ kako bi se dobili bolji zaključci o potrebnim promjenama kod poslovne analitike. Pomaže korisnicima u stvaranju različitih vizualizacija kao što su linijski grafikoni, kružni grafikoni, karte, heat mape itd. Na slici 3 prikazana je vizualizacija koje su pjesme poznatog benda „The Beatles“ bile popularne kroz godine. Sve navedene vizualizacije služe kao pomoć pri analizi postojećeg stanja i predviđanju budućih trendova.

³ Poslovna inteligencija – obuhvaća koncepte i tehnologije koje koriste poduzeća za analizu podataka i upravljanje poslovnim informacijama (Business Intelligence, 2023).



Slika 3 - Primjer vizualizacije u Tableau, vizualizaciju napravio: Tabelau Public korisnik Adam McCann, preuzeto sa: (TechTarget, 2019)

Tableau ima nekoliko bitnih značajki koje ga čine jednim od najpopularnijih alata u poslovnoj inteligenciji:

- omogućuje otkrivanje i analizu podataka koji korisnicima daju brze odgovore na postavljena pitanja ili nejasnoće,
- za korištenje Tableau-a, nije potrebno predznanje programiranja, što mu daje veliku prednost jer korisnici mogu bez iskustva koristiti softver.

Kako bi Tableau mogli koristiti svi korisnici (bilo početnici ili napredniji), softver je prilagođen za korištenje svima, odnosno postoji više vrsta proizvoda koji se nudi, a prikazani su u tablici 1.

Tablica 1 - Usporedba različitih vrsta Tableau-a, tablica napravljena na temelju: (Simplilearn, 2023)

Vrsta proizvoda	Verzija/Licenca	Karakteristike proizvoda
Tableau Desktop	Plaća se za korištenje	Pomaže u stvaranju izvješća, interaktivnih dashboard – a (nadzornih ploča) i priča pomoću različitih dijagrama i grafikona.
Tableau Server	Plaća se za korištenje	Najčešće se koristi u tvrtkama za dijeljenje radnih knjiga i grafikona stvorenih u Tableau Desktop između timova/kolega.
Tableau Online	Plaća se za korištenje	Ova verzija Tableau sofvera smještena je u oblaku (<i>engl. Cloud</i>) ⁴ , omogućuje korisnicima dijeljenje sadržaja

⁴ Pohrana u oblaku je model pohrane računalnih podataka u kojem se digitalni podaci pohranjuju u logičke skupove (Cloud Storage, 2023)

		i povezivanje s bazama podataka u oblaku.
Tableau Public	Besplatno za korištenje	Iako je besplatan za korištenje, korisnici nisu u mogućnosti radne knjige spremati lokalno. Radne knjige mogu se spremati u oblak i svaki korisnik ima pristup.
Tableau Reader	Besplatno za korištenje	Aplikacija za stolna računala koja služi korisnicima za čitanje i interakciju s izvješćima.
Tableau Mobile	Besplatno za korištenje	Omogućuje korisnicima pregled i interakciju s radnim knjigama i podacima s bilo kojeg mjesta u bilo koje vrijeme, mogućnost filtriranja podataka.

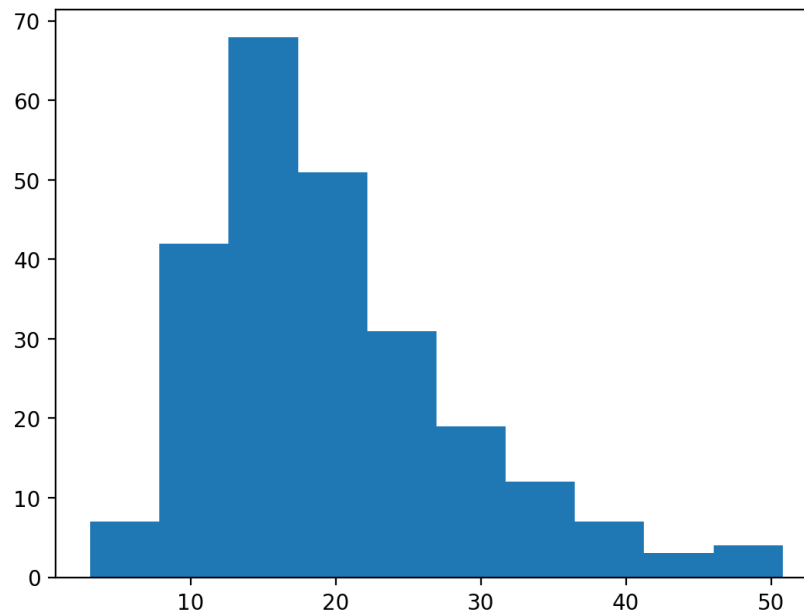
Kako bi korisnik koji iščitava podatke lakše shvatio rezultate provedenih analiza, podaci se vizualiziraju. Grafikonu u Pythonu mogu se dobiti kao kombinacije dvije biblioteke: *NumPy-a* i *Matplotlib-a*.

Biblioteka *NumPy* je biblioteka otvorenog koda (*engl. Open – source*) i vrlo je popularna jer uvelike poboljšava jednostavnost i performanse rada s višedimenzionalnim nizovima. *NumPy* je osnovni paket za znanstveno računalstvo u Pythonu. Pruža rad s višedimenzionalnim nizovima, raznim izvedenim objektima (npr. maskirani nizovi i matrice) i pomoć pri radu s „brzim operacijama“ na nizovima, uključujući matematičke i logičke operacije, sortiranje, osnovnu linearnu algebru, osnovne statističke operacije i još mnogo toga. Instalira se pomoću naredbe *pip install numpy*. Na slici 4 prikazan je *NumPy* niz.



Slika 4 - *NumPy* array, preuzeto s: (Developers, 2022)

Biblioteka *Matplotlib* koristi se za stvaranje raznovidnih 2D ili 3D vizualizacija u Pythonu. Vizualizacije mogu biti statične ili dinamične (animirane), a neke od vizualizacija koje biblioteka sadrži jesu stupčasti grafikon, histogram i violinski graf. Instaliranje je omogućeno naredbom *pip install matplotlib*. Na slici 5 Slika 5 prikazan je primjer histograma u *Matplotlib* – u.



Slika 5 - Primjer histograma u Matplotlib-u, preuzeto sa: (Matplotlib, 2020)

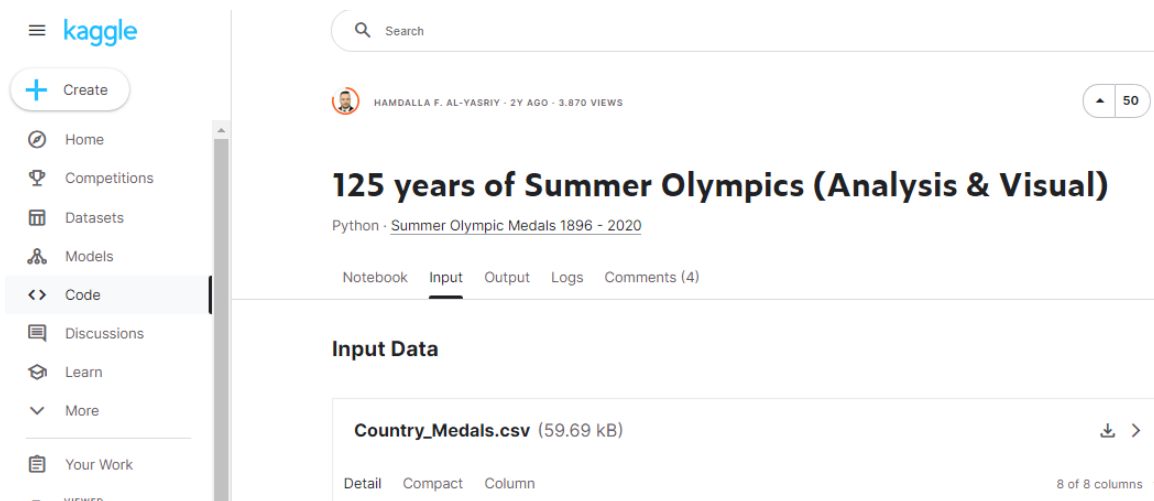
Kod vizualizacije podataka ključno je da su prikazi na grafikonima lako razumljivi korisniku koji iščitava grafove.

3. Opis i obrada skupa podataka

Prvi skup podataka koji ću obrađivati u ovom završnom radu je skup podataka koji je preuzet s *Kaggle* platforme. *Kaggle* je platforma za istraživanje znanosti o podacima i *online* zajednica znanstvenika koji se bave obradom podataka. Ona omogućuje korisnicima da objave i pronađu skup podataka koji im je potreban kod istraživanja, omogućava im izradu modela u okruženju znanosti o podacima te rad s drugim znanstvenicima o podacima i inženjerima strojnog učenja (Kaggle, 2023).

Ime prvog skupa podataka je „*Country_Medals.csv*“. Ovaj skup podataka pohranjen je kao podatkovni okvir. Podatkovni okvir je struktura podataka koja organizira podatke u tablični format s redovima i stupcima. Mogu se pohraniti u različite formate, kao npr. CSV⁵, Excel ili SQL baze podataka. Podaci koje navedeni skup sadrži vezan je uz osvojene medalje na ljetnim olimpijskim igrama po državama od 1896. godine do 2020. godine (preuzeto s: (125 years of Summer Olympics (Analysis & Visual), 2021)). Na slici 6 prikazana je stranica s koje su podaci preuzeti te ime datoteke koja se koristi pri obradi („*Country_Medals.csv*“).

⁵ Comma-separated values



Slika 6 - Preuzimanje prvog skupa podataka

Prikaz podatkovnog okvira može se vidjeti na slici 7:

	A	B	C	D	E	F	G	H
1	Year	Country_Code	Country_Name	Host_city	Host_country	Gold	Silver	Bronze
2	1932	(USA)	United States	Los Angeles	United States	41	32	30
3	1932	(ITA)	Italy	Los Angeles	United States	12	12	12
4	1932	(FRA)	France	Los Angeles	United States	10	5	4
5	1932	(SWE)	Sweden	Los Angeles	United States	9	5	9
6	1932	(JPN)	Japan	Los Angeles	United States	7	7	4
7	1932	(HUN)	Hungary	Los Angeles	United States	6	4	5
8	1932	(FIN)	Finland	Los Angeles	United States	5	8	12
9	1932	(GBR)	Great Britain	Los Angeles	United States	4	7	5
10	1932	(GER)	Germany	Los Angeles	United States	3	12	5
11	1932	(AUS)	Australia	Los Angeles	United States	3	1	1
12	1932	(ARG)	Argentina	Los Angeles	United States	3	1	0
13	1932	(CAN)	Canada	Los Angeles	United States	2	5	8
14	1932	(NED)	Netherlands	Los Angeles	United States	2	5	0
15	1932	(POL)	Poland	Los Angeles	United States	2	1	4
16	1932	(RSA)	South Africa	Los Angeles	United States	2	0	3
17	1932	(IRL)	Ireland	Los Angeles	United States	2	0	0
18	1932	(TCH)	Czechoslovakia	Los Angeles	United States	1	2	1
19	1932	(AUT)	Austria	Los Angeles	United States	1	1	3
20	1932	(IND)	India	Los Angeles	United States	1	0	0

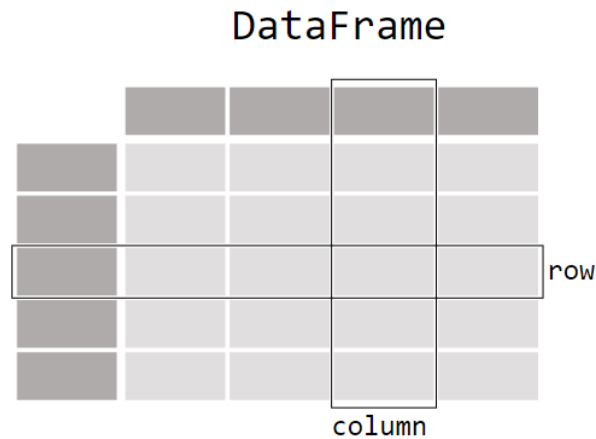
Slika 7 - Prikaz tablice u Excelu

Kao što vidimo, sadrži 8 stupaca:

- Year – godina kada su se održale olimpijske igre, od 1896. godine do 2020. godine,
- Country_Code – skraćeno ime države, njezina oznaka,
- Country_Name – ime države koja je sudjelovala u OI,
- Host_city – ime grada koji je bio domaćin za prethodno navedenu godinu,
- Host_country – ime države koja je bila domaćin za prethodno navedenu godinu,
- Gold – broj zlatnih medalja koja je određena država osvojila navedene godine u stupcu 'Year',
- Silver – broj srebrnih medalja koja je određena država osvojila navedene godine u stupcu 'Year',

- Bronze – broj brončanih medalja koja je određena država osvojila navedene godine u stupcu 'Year'.

Za rad i manipulaciju podatkovnim okvirima najčešće se koristi specijalizirana *Python* datoteka za podatkovne okvire: *Pandas DataFrame* čiji se prikaz može vidjeti na slici 8. *Pandas* se koristi pri obradi tabličnih podataka na način da pomaže pri istraživanju, čišćenju i obrađivanju podataka. *Pandas DataFrame* sastoji se od redaka i stupaca.



Slika 8 - Prikaz DataFrame - a, preuzeto s: (Getting Started, 2023)

3.1. Učitavanje u Pandas DataFrame

Podatkovni okvir učitani je u *Visual Studio Code* kako bi se s njime dalje mogla provoditi analiza i dobiti kvalitetni rezultati. Kako bi se započela analiza podatkovnog okvira, prvo se u novi *Python file* dodaju biblioteke koje su korištene za početak rada tj. za učitavanje originalne *.csv* datoteke u *Pandas DataFrame*. Način na koji se navedene naredbe izvode, može se vidjeti na slici 9:

```
1 import numpy as np
2 from pandas import read_csv
3 import pandas as pd
```

Slika 9 - Dodavanje početnih biblioteka

NumPy biblioteka se uobičajeno dodaje pod aliasom *np* zbog jednostavnosti korištenja. Isto vrijedi i za biblioteku *pandas* koja se dodaje pod aliasom *pd*. Pomoću funkcije *read_csv()* koja se nalazi u *pandas* biblioteci, pročitani je podatkovni okvir prikazan na slici 10. Biblioteka također služi kao pomoć pri ispisu podatkovnog okvira.

```
df_unsorted=pd.read_csv(r'C:\Users\korisnik\Desktop\3.godina\ZAVRSNI_RAD\Country_Medals.csv', sep=";")
print(df_unsorted)
```

Slika 10 - Učitavanje DataFrame-a

Putanja do datoteke navedena je pomoću *r("")*⁶, gdje se između navodnika navodi točna putanja, a *sep=";"* označava da su podaci u datoteci odvojeni točka-zarezom. Pomoću funkcije *print()* DataFrame pod imenom *df_unsorted* ispisan je na ekran te se njegov ispis prikazuje na slici 11. Kasnije će biti sortirani zbog lakšeg i intuitivnijeg korištenja. Na kraju učitavanja, nesortirani *Pandas DataFrame* izgleda ovako:

```
>>> print(df)
   Year Country_Code Country_Name Host_city Host_country Gold Silver Bronze
0   1932      (USA)  United States Los Angeles  United States    41     32     30
1   1932      (ITA)      Italy Los Angeles  United States    12     12     12
2   1932      (FRA)      France Los Angeles  United States    10      5      4
3   1932      (SWE)      Sweden Los Angeles  United States     9      5      9
4   1932      (JPN)      Japan  Los Angeles  United States     7      7      4
...   ...   ...   ...   ...   ...   ...   ...   ...
1339 1936      (MEX)      Mexico      Berlin      Germany      0      0      3
1340 1936      (BEL)      Belgium      Berlin      Germany      0      0      2
1341 1936      (AUS)      Australia      Berlin      Germany      0      0      1
1342 1936      (PHI)  Philippines      Berlin      Germany      0      0      1
1343 1936      (POR)      Portugal      Berlin      Germany      0      0      1

[1344 rows x 8 columns]
```

Slika 11 - Ispis nesortiranog DataFrame-a

⁶ Engl. read

U ispisu se ne prikazuje cijeli podatkovni okvir jer je prevelik, odnosno ima 1344 redaka. Ipak, prikazuju se nazivi svih 8 prethodno objašnjenih stupaca. Na primjer, vidi se kako su Sjedinjene Američke Države godine 1932. na ljetnim olimpijskim igrama, gdje je domaćin bio grad Los Angeles, osvojile 41 zlatnu medalju, 32 srebrne medalje i 30 brončanih medalja.

Zbog lakšeg snalaženja, podatkovni okvir sortiran je po godinama kada su se odvijale Olimpijske igre.

```
df = df_unsorted.sort_values(by="Year", ascending=False)
print(df)
```

Slika 12 - Funkcija sortiranja DataFrame-a

Slika 12 prikazuje funkciju pomoću koje se podatkovni okvir sortira. Novi podatkovni okvir nazvan je *df* i pomoću funkcije *sort_values(by="Year, ascending=False)* podaci su sortirani po stupcu *Year* koji se nalazi u originalnoj tablici i to uzlaznim redosljedom. Ispisan je na ekran funkcijom *print(df)*, pa ispis koji je prikazan na slici 13 izgleda ovako:

	Year	Country_Code	Country_Name	Host_city	Host_country	Gold	Silver	Bronze
182	1896	(GRE)	Greece	Athens	Greece	10	18	19
191	1896	(ZZX)	Mixed team	Athens	Greece	1	0	1
190	1896	(SUI)	Switzerland	Athens	Greece	1	2	0
189	1896	(DEN)	Denmark	Athens	Greece	1	2	3
188	1896	(AUS)	Australia	Athens	Greece	2	0	0
..
558	2020	(COL)	Colombia	Tokyo	Japan	0	4	1
560	2020	(DOM)	Dominican Republic	Tokyo	Japan	0	3	2
561	2020	(ARM)	Armenia	Tokyo	Japan	0	2	2
551	2020	(EST)	Estonia	Tokyo	Japan	1	0	1
559	2020	(AZE)	Azerbaijan	Tokyo	Japan	0	3	4

Slika 13 - Ispis sortiranog DataFrame-a

Možemo zaključiti da je podatkovni okvir poprilično intuitivan i da se podaci iz njega lako iščitavaju.

3.2. Korištene statističke metode

Kod ovog primjera može se vrlo jednostavno prikazati koja zemlja, od svih koje se nalaze u podatkovnom okviru, ima najviše osvojenih medalja, a koja ima najmanje. Kako bi to bilo moguće prikazati, potrebno je, zbog jednostavnosti, stvoriti novi stupac koji sadrži zbroj svih zlatnih, srebrnih i brončanih medalja za svaku godinu ljetnih OI za svaku zemlju koja je sudjelovala. Na slici 14 prikazano je stvaranje novog stupca.

```
df["All_Medals"]=df["Gold"]+df["Silver"]+df["Bronze"]
print(df)
```

Slika 14 - Dodavanje novog stupca "All_Medals"

Novi stupac nazvan je *All_Medals* i nastao je kao zbroj svih zlatnih, srebrnih i brončanih medalja za svaku zemlju. Iz podatkovnog okvira *df* zbrojena su tri stupca, *Gold*, *Silver* i *Bronze* te se dodaju u novi stupac, ali u isti podatkovni okvir. Ispis je napravljen pomoću funkcije *print(df)* te je prikazan na slici 15:

```
>>> df["All_Medals"]=df["Gold"]+df["Silver"]+df["Bronze"]
>>> print(df)
   Year Country_Code Country_Name Host_city Host_country Gold Silver Bronze All_Medals
182  1896      (GRE)      Greece Athens Greece      10      18      19         47
191  1896      (ZZX)  Mixed team Athens Greece       1       0       1          2
190  1896      (SUI)  Switzerland Athens Greece       1       2       0          3
189  1896      (DEN)      Denmark Athens Greece       1       2       3          6
188  1896      (AUS)      Australia Athens Greece       2       0       0          2
..   ...      ...      ...      ...      ...      ...      ...      ...
558  2020      (COL)      Colombia Tokyo Japan       0       4       1          5
560  2020      (DOM) Dominican Republic Tokyo Japan       0       3       2          5
561  2020      (ARM)      Armenia Tokyo Japan       0       2       2          4
551  2020      (EST)      Estonia Tokyo Japan       1       0       1          2
559  2020      (AZE)      Azerbaijan Tokyo Japan       0       3       4          7

[1344 rows x 9 columns]
```

Slika 15 - Ispis novog stupca "All_Medals"

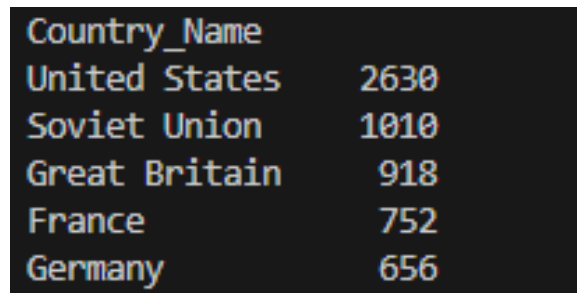
Kao što se može vidjeti, novi stupac je dodan i sada podatkovni okvir ima isto redaka kao i prije, 1344, ali jedan stupac više. Novi stupac je napravljen radi jednostavnosti, a s ciljem da se prikaže zemlja koja ima najviše osvojenih medalja i zemlja koja ima najmanje osvojenih medalja.

Sljedeći isječak koda na slici 16 pomaže pri prikazu zemlji s najviše osvojenih medalja:

```
medalje_max=df.groupby("Country_Name")["All_Medals"].sum().sort_values(ascending=False).head()
print(medalje_max)
```

Slika 16 - Prikaz koda za pet država s najviše medalja

Stvara se varijabla `medalje_max` koja nam predstavlja varijablu država koje su osvojile najviše medalja. Dio linije `df.groupby(„Country_Name“).[„All_Medals“].sum()` grupira nam podatkovni okvir po stupcu `Country_Name`, te se sve medalje iz stupca `All_Medals` zbrajaju za svaku državu posebno pomoću funkcije `sum()`. Funkcija `sort_values(ascending=False)` sortira vrijednosti od gore prema dolje, odnosno od države koja ima najviše medalja do države koja ima najmanje medalja. Zbog nepotrebnog ispisivanja cijelog podatkovnog okvira, koristila se funkcija `head()`⁷ kako bi se ispisalo prvih 5 država u podatkovnom okviru po broju medalja. Funkcijom `print(medalje_max)` podatkovni okvir ispisan je na ekran, a prikazan je na slici 17:



Country_Name	
United States	2630
Soviet Union	1010
Great Britain	918
France	752
Germany	656

Slika 17 - Prikaz ispisa pet zemalja s najviše medalja

Sam ispis zemlje s najviše medalja događa se u sljedećem isječku koda koji je prikazan na slici 18:

```
zemlja_max=medalje_max.idxmax("Country_Name")
broj_max=medalje_max.max()
print ("Zemlja s najviše medalja je", zemlja_max, "i ima ", broj_max, "medalju.")
```

Slika 18 – Kod sa ispis zemlje s najviše medalja

Varijabla `zemlja_max` pohranjuje ime zemlje koja ima najviše osvojenih medalja. Za pohraniti ime u varijablu `zemlja_max`, koristi se funkcija `idxmax()`. Funkcija može sadržavati tri atributa:

- `axis` – ako je `axis=0` ili `index`, traži se najveća vrijednost po retcima, a ako je `axis=1` ili `column` traži se najveća vrijednost po stupcima, `default` vrijednost je 0.
- `skipna` – ovaj se atribut koristi ako se pojavljuju nedostajuće vrijednosti (NaN). Ako je `skipna=True` ignoriraju se sve nedostajuće vrijednosti, a ako je `skipna=False` vraća se NaN ako je bilo koji od elemenata NaN; `default` vrijednost je `True`.
- `numeric_only` – atribut uključuje samo numeričke vrijednosti, `default` vrijednost je `False`

U gornjem primjeru,, funkcija vraća maksimalnu vrijednost tj. ime države (indeks) iz podatkovnog okvira `medalje_max`. Funkcija `max()` vraća maksimalan broj medalja iz podatkovnog okvira `medalje` te sprema broj medalja u varijablu `broj_max`. Zatim se tražena zemlja i broj medalja koje je osvojila kroz sve godine ljetnih OI (od 1896. godine do 2020. godine) ispisuju na ekran, a ispis je prikazan na slici 19:

⁷ `Default` vrijednost je 5, pa nije potrebno specijalno između zagrada navoditi broj 5

```
Zemlja s najviše medalja je United States i ima 2630 medalju.
```

Slika 19 - Ispis zemlje s najviše medalja

Na isti način, na slici 20, napravljen je ispis zemalja koje na OI imaju najmanje osvojenih medalja:

```
medalje_min=df.groupby("Country_Name")["All_Medals"].sum().sort_values(ascending=False).tail()
print(medalje_min)
```

Slika 20 - Prikaz koda za pet država s najmanje medalja

Stvara se varijabla *medalje_min* koja nam predstavlja varijablu država koje su osvojile najmanje medalja. Ponovo se koristi funkcija *groupby()* za grupiranje po stupcu *Country_Name* i funkcija *sum()* za zbrajanje svih medalja. Također se koristi i funkcija *sort_values()* koja sortira vrijednosti od gore prema dolje. Jedina razlika u odnosu na prethodni isječak koda je u funkciji *tail()*⁸ koja služi za ispisivanje podatkovnog okvira od kraja datoteke. Funkcijom *print(medalje_min)*, podatkovni okvir prikazan na slici 21 se ispisuje na ekran:

```
Country_Name
Gabon          1
Iraq           1
Guatemala     1
Guyana        1
Paraguay      1
```

Slika 21 - Prikaz ispisa pet zemalja s najmanje medalja

Pošto postoji više zemalja koje su osvojile samo jednu medalju, nije logično da u ispisu postoji samo jedna zemlja koja ima jednu medalju, već mora postojati lista zemalja koje su kroz sve godine ljetnih olimpijskih igara osvojile samo jednu medalju.

Za ispis svih zemalja sa jednom osvojenom medaljom služi sljedeći isječak koda na slici 22:

```
zemlja_min = df.groupby("Country_Name")["All_Medals"].sum()
zemlje_jed_med = zemlja_min[zemlja_min == 1]
print("Zemlje sa samo 1 medaljom su:", ", ".join(zemlje_jed_med.index), ".")
```

Slika 22 - Lista svih zemalja sa jednom osvojenom medaljom

⁸ Default vrijednost je 5, pa nije potrebno specijalno između zagrada navoditi broj 5

U varijablu `zemlja_min` spremaju se sve zemlje koje su sudjelovale u ljetnim OI i funkcijom `groupby()` grupiraju se po imenima država (`Country_Name`) te se sve osvojene medalje po državama zbrajaju funkcijom `sum()` kako u ispisu ne bismo imali više redaka s istom državom. U varijablu `zemlje_jed_med` spremaju se sve zemlje koje su na ljetnim OI osvojile jednu medalju. To se provjerava uvjetom `zemlja_min==1`.

Ovdje se koristi i funkcija `join()` koja služi za spajanje stupaca iz različitih podatkovnih okvira, a stupci se mogu spajati preko indeksa ili preko ključa u stupcu. Funkcija `join()` sadrži nekoliko parametara:

- `other` – `DataFrame`, `Series`⁹ ili lista koja sadrži kombinaciju istih,
- `on` – `str`, `list of str`, `or array-like`, `optional`,
- `dtype` – `str`, `numpy.dtype`, `or ExtensionDtype`, `optional`,
- `name` – `Hashable`, `default None`,
- `copy` – `bool`, `default false`.

U gornjem primjeru, spajanje stupaca događa se preko indeksa i u listu se zapisuju sve zemlje koje su osvojile jednu medalju na ljetnim OI, odvojene zarezom. Dobiveni ispis prikazan je na slici 23:

```
Zemlje sa samo 1 medaljom su: Barbados, Burkina Faso, Ceylon, Cyprus, Djibouti, Eritrea, Gabon, Guatemala, Guyana, Iraq, Macedonia, Mauritius, Montenegro, Netherlands Antilles, North Macedonia, Paraguay, Samoa, Senegal, Sri Lanka, Sudan, Togo, Tonga, Turkmenistan, Virgin Islands .
```

Slika 23 - Ispis zemalja sa jednom medaljom

Zbog znatiželje, napravljen je i upit za zemlje koje nisu osvojile ni jednu medalju. Upit je prikazan na slici 24:

```
zemlja_0=zemlja_min[zemlja_min == 0]  
print("Zemlje koje nisu osvojile ni jednu medalju su:", ", ".join(zemlja_0.index), ".")
```

Slika 24 - Upit za zemlje koje nemaju osvojenu ni jednu medalju

U varijablu `zemlje_0` spremaju se sve zemlje koje na ljetnim OI nisu osvojile ni jednu medalju. To se provjerava uvjetom `zemlja_min==0`. Kod ispisa se ponovo koristi funkcija `join()`, ali u ovom slučaju ne postoji ni jedna zemlja koja nije osvojila ni jednu medalju. Ispis je prikazan na slici 25:

```
Zemlje koje nisu osvojile ni jednu medalju su: .
```

Slika 25 - Ispis zemalja koje nemaju ni jednu medalju

Nadalje, napravljen je upit prikazan na slici 26 nad podatkovnim okvirom `df` koji na ekran ispisuje sve rezultate Hrvatske na svim ljetnim OI.

⁹ Jednodimenzionalni niz s oznakama osi

```
hrvatska=df.loc[df['Country_Name'] == 'Croatia'].drop(columns=["Country_Code", "Host_country"])
print (hrvatska)
```

Slika 26 - Kod za ispis rezultata Hrvatske

Koristi se `loc()` metoda kako bi se pristupilo grupi redaka i/ili stupaca preko specijalne oznake ili boolean array – a. Grupe redaka i/ili stupaca spremaju se u varijablu `hrvatska`. U ovom slučaju, može se uočiti da je to uvjet `['Country_Name'] == 'Croatia'` te metoda pronalazi sve retke i/ili stupce u tablici gdje je ime države 'Croatia' te ispisuje sve povezane retke i stupce s tom specijalnom oznakom. Kod ispisa se koristila metoda `drop(columns=[])` za micanje određenih stupaca iz ispisa zbog jasnijeg pregleda konačnih rezultata. Nakon micanja stupaca `Country_Code` i `Host_Country` iz ispisa, izgled podatkovnog okvira prikazan je na slici 27:

	Year	Country_Name	Host_city	Gold	Silver	Bronze	All_Medals
518	2020	Croatia	Tokyo	3	3	2	8
602	2016	Croatia	Rio de Janeiro	5	3	2	10
697	2012	Croatia	London	3	1	2	6
816	2008	Croatia	Beijing	0	2	3	5
888	2004	Croatia	Athens	1	2	2	5
967	2000	Croatia	Sydney	1	0	1	2
1044	1996	Croatia	Atlanta	1	1	0	2
1122	1992	Croatia	Barcelona	0	1	2	3

Slika 27 - Rezultati Hrvatske na ljetnim OI

Zbog jasnoće prikaza i razumijevanja rezultata, napravljen je grafikon koji prikazuje Hrvatsku kroz sve godine kada je sudjelovala u ljetnim OI. Da bi se grafikon mogao stvoriti, potrebno je na početak datoteke uvesti biblioteku `matplotlib` i modul `pyplot` za crtanje grafova i vizualizaciju podataka. To se radi pomoću naredbe `import matplotlib.pyplot as plt`. Nakon toga piše se kod.

Sljedeći isječak koda (slika 28) izradit će graf prikazan na slici 29:

```
god=hrvatska["Year"]
podaci=hrvatska["All_Medals"]
plt.figure(figsize=(12, 6))
plt.xticks([1992, 1996, 2000, 2004, 2008, 2012, 2016, 2020])
plt.plot(god, podaci, marker="o", linestyle="-", color="red")
plt.title("Hrvatska kroz godine")
plt.xlabel("Godina")
plt.ylabel("Broj osvojenih medalja")
plt.grid(True)
plt.show()
```

Slika 28 – Grafikon napravljen pomoću Matplotlib – a

U varijable `god` i `podaci` izdvajaju se stupci iz podatkovnog okvira `hrvatska` koji je korišten u prethodnom primjeru. Iz njega se izdvajaju stupci `Year` i `All_Medals` kako bi se na grafu mogle prikazati godine i broj osvojenih medalja za Hrvatsku. Linija `plt.figure(figsize=12,6)` postavlja veličinu grafa na 12x6 inča. Pomoću atributa `xticks` postavljaju se oznake na x – osi. U ovom slučaju to su sve godine u kojima je Hrvatska sudjelovala na ljetnim OI. Funkcija `.plot()` koristi se za vizualizaciju podataka na

grafu. Na x – osi će se prikazivati godine, a na y – osi broj osvojenih medalja. Opcija *marker="o"* definira označavanje svake točke na grafu, u ovom slučaju to će biti kružni markeri, tj. pune točke. Pomoću *linestyle="-"* definira se neprekidna linija koja će služiti za povezivanje svih točaka. Funkcijom *plt.title()* dao se naslov grafu „Hrvatska kroz godine“. Funkcijama *plt.xlabel()* i *plt.ylabel()* dala su se imena x i y osima. Pomoću linije *plt.grid(True)* prikazala se je mreža na grafu, a linijom *plt.show()* cijeli grafikon prikazao se kao slika. Dobiveni grafikon izgleda ovako (slika 29):



Slika 29 - Medalje za Hrvatsku kroz godine

3.3. Ulaganja u sport i korelacija

Ulaganja u sport

Kako bi se pokazala povezanost između ulaganja u sport i uspjeha tj. utječu li financijska ulaganja u reprezentacije na njihove rezultate, iz prijašnje tablice „*Country_Medals.csv*“ izdvojiti će se nekoliko Europskih zemalja i u novu tablicu će se dodati stupac „*Investments_per_Inhabitant*“ koji će predstavljati ulaganje u pojedinog sportaša za nekoliko Europskih država.

Pošto je ovdje najbitnija povezanost između ulaganja u sport i uspjeha, neki stupci biti će izbačeni iz nove tablice zbog jednostavnosti i lakšeg razumijevanja rezultata. Tablica je napravljena na temelju iščitavanja iz grafikona s platforme Eurostat (How much do governments spend on recreation and sport?, 2020). Podaci o ulaganjima su iz 2018. godine, a tablica je prikazana na slici 30:

1	Year	Country_Name	Gold	Silver	Bronze	All_Medals	Investments_per_Inhabitant(€)
2	2016	Germany	17	10	15	42	101
3	2016	France	10	18	14	42	204
4	2016	Netherlands	8	7	4	19	202
5	2016	Hungary	8	3	4	15	120
6	2016	Spain	7	4	6	17	101
7	2016	Croatia	5	3	2	10	39
8	2016	Sweden	2	6	3	11	254
9	2016	Finland	0	0	1	1	226
10	2016	Denmark	2	6	7	15	201
11	2016	Bulgaria	0	1	2	3	11
12	2016	Slovakia	2	2	0	4	31
13	2016	Lithuania	0	1	3	4	35

Slika 30 - Tablica sa podacima o ulaganjima

Tablica ima 7 stupaca i 12 redaka u kojima se nalaze podaci iz tablice „*Country_Medals.csv*“ za nekoliko država koje se istovremeno nalaze u Europi i u Europskoj Uniji. Neki stupci su izbačeni zbog lakšeg razumijevanja rezultata. Dodan je stupac „*Investments_per_Inhabitant*“ koji nam govori koliko je svaka država 2018. godine uložila u svakog svog sportaša (u Eurima).

Korelacija

Korelacija predstavlja statističku mjeru koja se može primijeniti samo na numeričke varijable te opisuje stupanj povezanosti između te dvije varijable (Rahman, Ozair, Khan, & Sonkar, 2023). Drugim riječima, korelacija govori postoji li odnos između dvije varijable. Ako se varijable kreću u istim smjerovima, onda korelacija može biti najviše 1 (savršena pozitivna korelacija). To znači da postoji jaka pozitivna korelacija tj. povećanje jedne varijable obično se povezuje s povećanjem druge varijable. Ako se varijable razlikuju, onda nam korelacija iznosi najmanje -1 (savršena negativna korelacija). Samim time, govori se o jakoj negativnoj korelaciji tj. smanjenje jedne varijable obično utječe na povećanje druge varijable i obratno. Time zaključujemo da nam je raspon korelacije od -1 do 1, a što je iznos korelacije bliže ili jednak nuli, zaključujemo da je korelacija slaba ili uopće ne postoji (Rahman, Ozair, Khan, & Sonkar, 2023). Koeficijent korelacije preporučljivo je izračunavati samo kod onih varijabli gdje je varijanca konačna što znači da je broj podataka u uzorku prebrojiv (Rahman, Ozair, Khan, & Sonkar, 2023). Dva najpoznatija načina za izračunavanje korelacije su: Pearsonov korelacijski koeficijent i Spearmanov korelacijski koeficijent.

3.3.1. Pearsonov korelacijski koeficijent

Pearsonov koeficijent korelacije upotrebljava se kada postoji linearna povezanost između promatranih varijabli (Rahman, Ozair, Khan, & Sonkar, 2023). Pearsonov koeficijent označava se malim grčkim slovom „ ρ “ i njegova vrijednost kreće se od -1 do 1. Što je vrijednost korelacijskog koeficijenta bliža ± 1 , jača je korelacija između varijabli. Tablica 2 predstavlja nam okvirne koeficijente različitih kategorija korelacije:

Tablica 2 - Kategorizacija snage korelacije u odnosu na korelacijski koeficijent, tablica napravljena na temelju: (Rahman, Ozair, Khan, & Sonkar, 2023, str. 184)

Korelacija	Pearsonov korelacijski koeficijent
Savršena	$r=\pm 1$
Vrlo visoka	$r=\pm 0.90-1$
Visoka	$r=\pm 0.70-0.90$
Srednja	$r=\pm 0.50-0.70$
Niska	$r=\pm 0.30-0.50$
Vrlo niska	$r=\pm 0.10-0.30$
Nepostojeća	$r=0$

3.3.2. Spearmanov korelacijski koeficijent

Spearmanov korelacijski koeficijent označen sa „ r “ koristi se kod računanja korelacije kod podataka s ekstremnim vrijednostima (outlieri)¹⁰ (Rahman, Ozair, Khan, & Sonkar, 2023). Distribucija kod takvih podataka zove se nenormalna distribucija, a oblik povezanosti između varijabli je sporedan (npr. linearna povezanost koja je uvjet za Pearsonov korelacijski koeficijent). On se koristi kada nije moguće koristiti Pearsonov koeficijent korelacije npr. u slučajevima kada ne postoji linearna povezanost među varijablama, a teško je ili skoro nemoguće primijeniti potrebnu transformaciju kojom bi se povezanost između varijabli prevela u linearnu. Što se koeficijenata tiče, Spearmanov koeficijent ima istu kategorizaciju snage korelacije u odnosu na korelacijski koeficijent pa za njega, isto kao i za Pearsonov koeficijent, vrijedi tablica 2.

3.3.3. Izračun i prikaz korelacije u Pythonu

Za novoizrađenu tablicu, izračunat će se korelacija između svih varijabli te će se ona vizualizirati pomoću Seaborn¹¹ u Pythonu. Prvo što je potrebno napraviti je uvođenje raznih biblioteka (slika 31) koje su potrebne za izračunavanje korelacije i prikaz odgovarajuće vizualizacije.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Slika 31 - Dodavanje odgovarajućih biblioteka u novi Python File

Jedina nova biblioteka koja se koristi je *Seaborn* i koristi se u Pythonu kao *sns* zbog lakšeg korištenja.

¹⁰ Outlieari predstavljaju vrijednosti varijabli koje u velikoj mjeri odstupaju od ostalih. (Korelacija, 2023)

¹¹ Seaborn je Python biblioteka za vizualizaciju podataka temeljena na matplotlibu. (Seaborn, 2022)

```
df=pd.read_csv(r'C:\Users\korisnik\Desktop\3.godina\ZAVRSNI_RAD\Ulaganja_po_drzavama.csv')
df=df.rename(columns={"Investments_per_Inhabitant(€)": "Investments"})
```

Slika 32 - Učitavanje u DataFrame

Na slici 32 prikazan je isječku koda gdje se .csv datoteka učitava u Pandas DataFrame pod imenom *df*. Ovdje se, zbog jednostavnosti korištenja DataFrame-a, koristi metoda za preimenovanje stupaca *.rename()* pomoću koje se stupac *Investments_per_Inhabitant(€)* preimenuje u *Investments*.

```
korelacija=df.corr()
print (korelacija)
plt.figure(figsize=(12,10))
sns.heatmap (data = korelacija, annot = True, linewidth=5, cmap=sns.color_palette("crest", as_cmap=True))
sns.set (font_scale = 1.0)
plt.show()
```

Slika 33 - Izračunavanje i prikaz korelacije pomoću HeatMape

Korelacija se na slici 33 izračunava pomoću metode *.corr()* te se izračunata korelacija sprema u varijablu *korelacija*. Veličina grafa postavlja se na 12x10 inča, a vizualizacija koja je odabrana za korištenje je heatmapa. Koristi se Seaborn za stvaranje heatmapa pomoću *sns.heatmap()* metode. Ona prikazuje korelaciju svih parova stupaca iz DataFrame – a. Dio metode *annot=True* dodaje tekst tj. brojke na grafikon kako bi bila prikazana točna vrijednost korelacije između stupaca. *Linewidth=5* postavlja širinu između linija heatmapa na 5 piksela, a *color_pallette(„crest“)* postavlja paletu boja na tamne i svjetlije nijanse plave boje kako bi jačina korelacije bila bolje uočljiva. Tamnijom bojom prikazana je jača korelacija, a svjetlijom slabija. Veličina fonta je 1.0. Dobivena heatmapa prikazana je na slici 34.



Slika 34 - Vizualizacija korelacije na Heatmapi

Da je rezultat korelacije između ulaganja i broja osvojenih medalja bio skoro 1, to bi navodilo na jaku pozitivnu korelaciju između tih dviju varijabli. Drugim riječima, to bi značilo da zemlje koje ulažu više

financijskih resursa u sportske aktivnosti obično postižu bolje rezultate na sportskim natjecanjima (u ovom slučaju više medalja na Olimpijskim igrama). To bi moglo ukazivati na to da financijska podrška, kao što su ulaganja u infrastrukturu, treninge sportaša i sportske programe, ima značajan utjecaj na postizanje uspjeha na sportskom području. Međutim, iz heatmapa se može očitati da dobiveni rezultat iz obrađenih skupova podataka ne ukazuje na to. Korelacija između stupca ulaganja i svih medalja iznosi 0.27, što znači da je korelacija mala, tj. prema Tablica 2, korelacija je vrlo niska, što znači da ulaganja u sport i uspjeh u sportu mogu biti povezani, ali to nije jedini faktor koji utječe na sportski uspjeh.

3.4. Metode vizualizacije

U ovom će poglavlju biti objašnjene neke metode vizualizacija koje postoje u Tableau. Naravno, neće biti moguće obuhvatiti sve metode vizualizacije jer ih ima mnogo, ali one koje se učestalije koriste i koje su važne za konkretne primjere u ovom radu, bit će detaljnije objašnjene.

Vizualizacije u Tableau

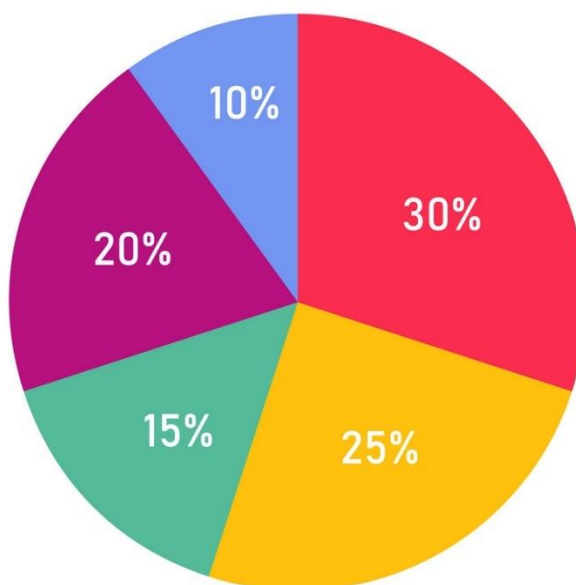
Kao što je već spomenuto u poglavlju 2.1. Tableau, postoje različite vrste grafikona i vizualizacija s istima. Grafikon je vizualna reprezentacija u obliku dijagrama, mape ili tablice (A Guide To Charts, 2023). Pomoću grafikona se napravljene analize nad podacima prikazuju u vizualnim oblicima. Na primjer, na grafovima se može prikazati usporedba podataka ili distribucija podataka. Prije nego li se odluči koja će se vrsta grafikona koristiti, podaci se moraju proučiti te se mora zaključiti što se ustvari želi pokazati publici. Prednost kod Tableau – a je da, kada se podaci koji se vizualiziraju ubace, softver ponudi prikladne grafove koji odgovaraju podacima. Karakteristike koje su preporučljive da grafikon ima jesu: dosljednost, jednostavnost, korištenje adekvatnih oznaka i fontova, korištenje boja (umjereno), jasan stil, relevantnost te razumijevanje publike.

Grafikoni koji se učestalije koriste kod Tableau – a su:

1. TORTNI GRAFIKON (*engl. Pie Chart*)

Tortni grafikon, kao što mu samo ime nalaže, koristi krug kao prikazivanje cjeline. Njegovi dijelovi se organiziraju kao cjelina, tj. jedna „kriška“ predstavlja specifičnu kategoriju. Kada se kategorije zbroje, one zajedno čine jednu cjelinu. Svaka od tih kategorija predstavlja neki udio vrijednosti i obično svaka kategorija predstavlja neki postotak grafikona.

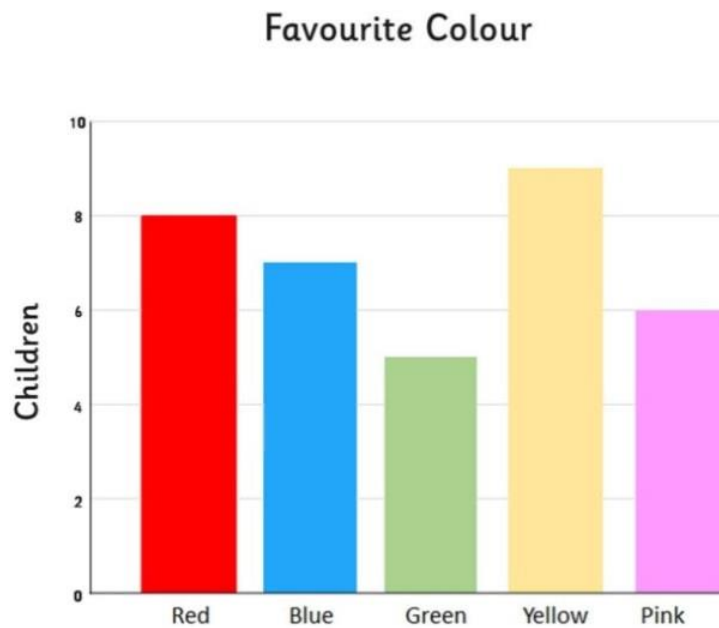
Za lakše razumijevanje tortnog grafikona, poželjno je kategorije organizirati po redoslijedu, odnosno od većeg dijela prema manjem ili obratno. Na taj je način korisniku najlakše razumjeti grafikon, bez dodatnog pojašnjavanja pomoću legende. Tortni grafikon prikazan je na slici 35:



Slika 35 - Primjer tortnog grafikona, preuzeto s (*Percentage pie chart for finance report, 2023*)

2. STUPČASTI GRAFIKON (engl. Bar Chart)

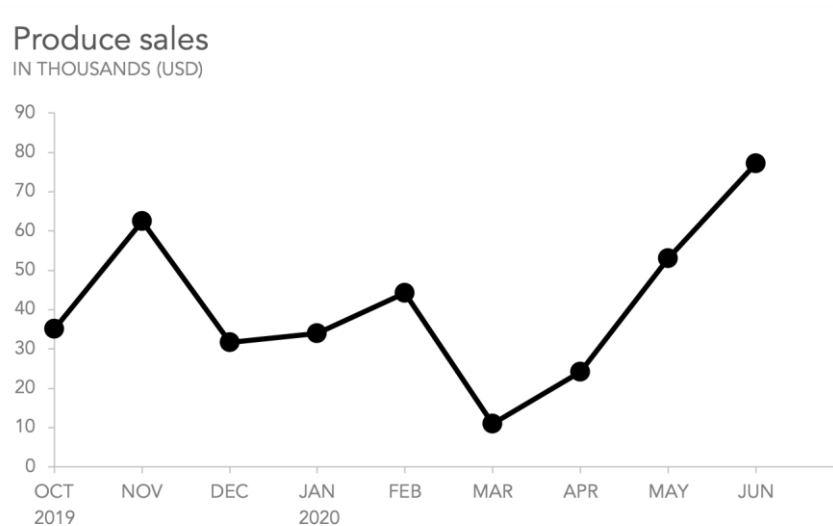
Stupčasti grafikon je vrsta grafikona koja omogućuje prikazivanje numeričkih varijabli, npr. cijelih brojeva i postotaka, a svaki stupac predstavlja vrijednost svake numeričke varijable. Vrijednosti se mogu prikazivati vodoravno (na x-osi) ili okomito (na y-osi). Prikaz će ovisiti o tome o kakvim se podacima radi. Primjer prikaza podataka sa stupčastim grafikonom (slika 36) je prikaz broja prodanih artikala unutar trgovine u nekom vremenskom intervalu.



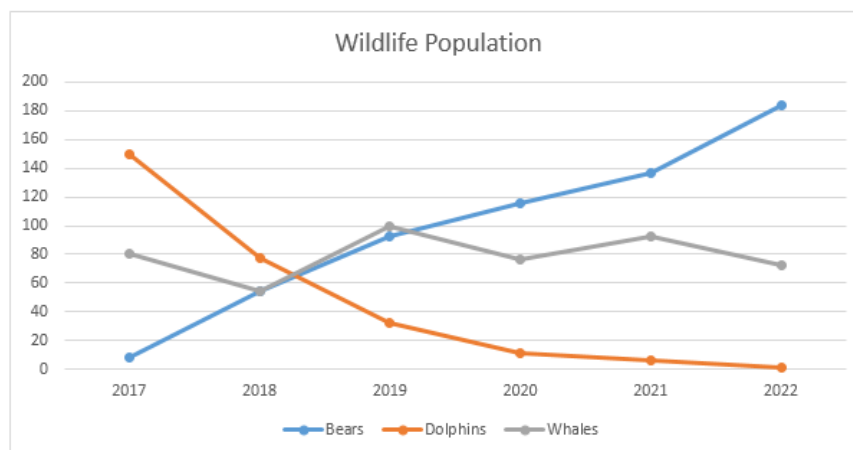
Slika 36 - Primjer stupičastog grafikona, preuzeto sa: (What is a Bar Chart?, 2022)

3. LINIJSKI GRAFIKON

Linijski grafikon je tip grafikona koji povezuje niz točaka, gdje svaka predstavlja neki podatak, u liniju. Najčešće predstavlja sekvencijalne vrijednosti, gdje vodoravna os (x – os) predstavlja slijedno kretanje vrijednosti, a iz okomite osi (y – os) se čitaju odgovarajuće vrijednosti. Većinom se koristi za prikazivanje podataka tokom vremena. U tom slučaju, x – os predstavlja dimenziju vremena, a y – os neku vrijednost (npr. novac, količinu proizvoda..). Na njemu se može prikazati jedna ili više linija, a prikaz ovisi o analizi koja se provodi i koju grafikon podržava. Analiza koja se najčešće provodi je analiza promjene trendova tokom vremena. To se vizualizira linijskim grafikonom sa samo jednom linijom (Slika 37), gdje smjer i oblik linije predstavljaju promjene. Ako linija na grafikonu raste s lijeva na desno, znači da grafikon označava rastući trend, a ako pada s lijeva na desno znači da je trend u silaznoj putanji. Postoji i linijski grafikon s više linija, gdje svaka linija predstavlja jedan dio istog područja koje se analizira (slika 38).



Slika 37 - Linijski grafikon s jednom linijom, preuzeto s: (Cisneros, 2020)

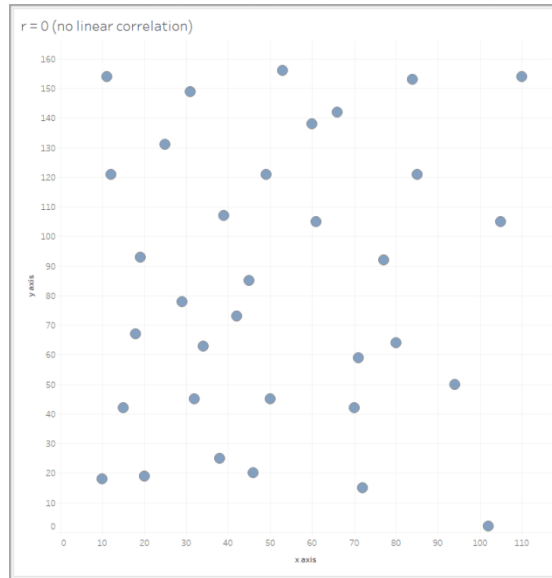


Slika 38 - Linijski grafikon s više linija, preuzeto s: (Line Chart, 2023)

4. DIJAGRAM RASPRŠENOSTI (engl. Scatterplot)

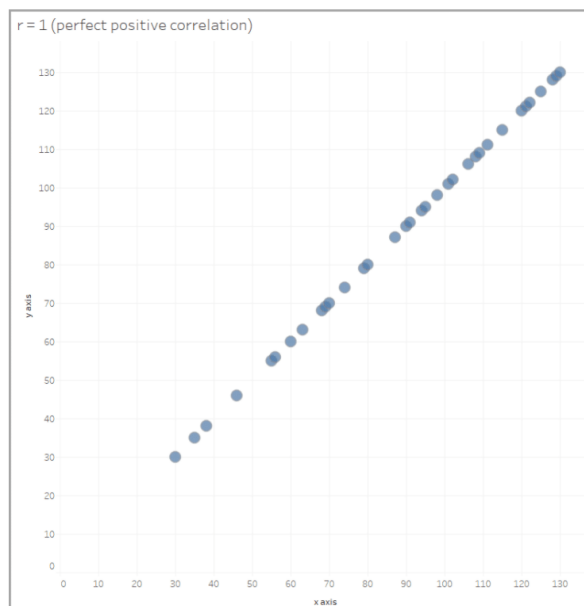
Dijagram raspršenosti prikazuje točke s podacima na grafikonu u kojoj se sijeku dvije mjere. Njima se olakšava analiza odnosa između dva broja jer su sve točke prikazane na istom prikazu. Baš kao i linijski grafikoni, dijagrami raspršenosti koriste Kartezijev koordinatni sustav za vizualizaciju podataka. Pomoću koordinata se iscrtavaju točke na karti i time se olakšava čitanje složenih podataka, jer se točke iscrtavaju na presjeku dviju koordinata (x i y). Što se koordinata tiče, x – os je uvijek nezavisna, a y – os je zavisna koordinata. Cilj dijagrama raspršenosti je prikazati kako promjene u nezavisnoj varijabli utječu na promjene u zavisnoj varijabli. Ovdje se može ispitati i prikazati korelacija između varijabli. Postoje tri primjera iščitavanja rezultata grafa kada je korelacija u pitanju:

- Iz grafa prikazanog na slici 39 može se vidjeti kako je korelacija vrlo slaba ili ne postoji, odnosno točke su raspršene po cijelom grafu.



Slika 39 - Korelacija ne postoji na dijagramu raspršenosti, preuzeto s: (A Guide To Charts, 2023)

- Graf na slici 40 predstavlja savršenu pozitivnu korelaciju ($r=1$) pomoću dijagrama raspršenosti. Ako su varijable proporcionalne¹², onda dobivamo savršenu pozitivnu korelaciju.

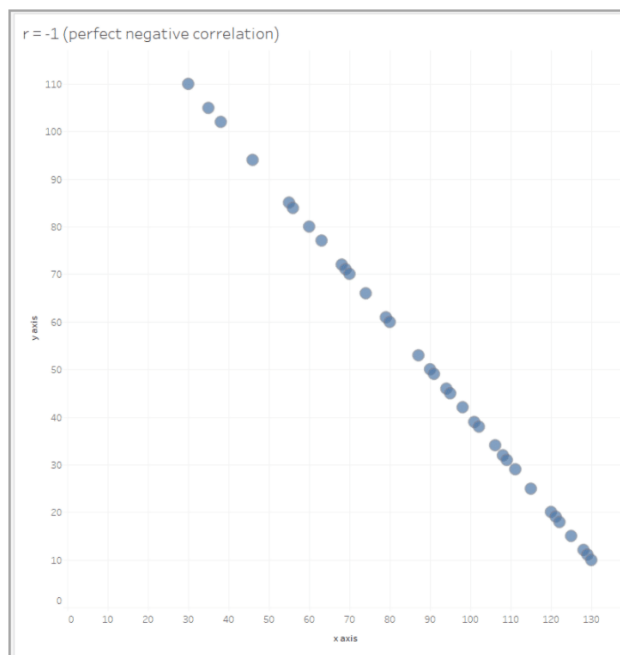


Slika 40 - Savršena pozitivna korelacija na dijagramu raspršenosti, preuzeto s: (A Guide To Charts, 2023)

- Graf na slici 41 predstavlja savršenu negativnu korelaciju ($r=-1$) pomoću dijagrama raspršenosti. Ako su varijable obrnuto proporcionalne¹³, onda dobivamo savršenu negativnu korelaciju.

¹² S rastom jedne varijable, raste i druga varijabla.

¹³ S padom jedne varijable raste druga varijabla i obrnuto.

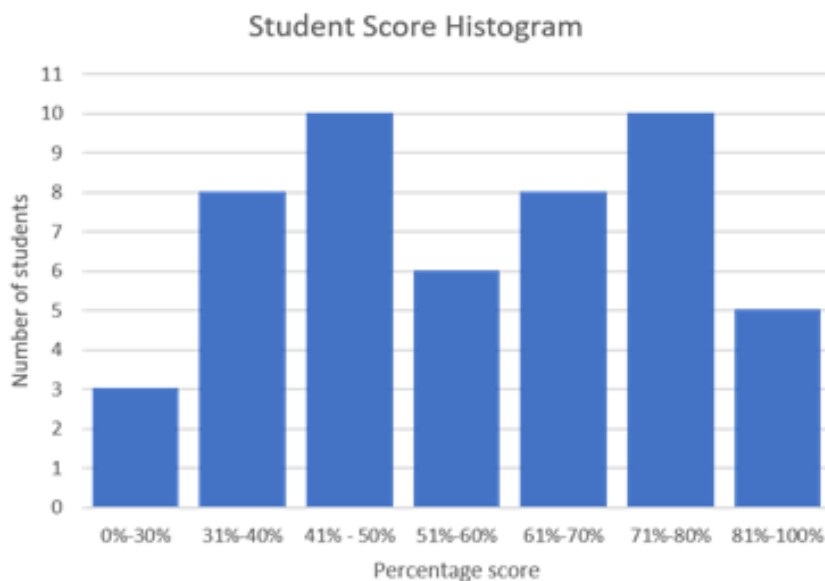


Slika 41 - Savršena negativna korelacija na dijagramu raspršenosti, preuzeto s: (A Guide To Charts, 2023)

5. HISTOGRAM (engl. Histogram)

Histogram (slika 42) predstavlja varijantu stupčastog grafikona te pruža način za prikazivanje distribucije podataka. Prati različite vrijednosti koje su dio nekog skupa podataka, kao niz povezanih stupaca. Vrijednosti su podijeljene u stupce ili skupine koje označavaju raspon vrijednosti. Svaki podatak se pridruži određenoj skupini te ih se slaže jedna do druge.

Najbolje su prikazani kada sadrže kontinuirane numeričke podatke, ali ne podržavaju usporedbe između dvije ili više kategorija, kao npr. stupčasti grafikoni. Dobar primjer histograma može se napraviti od podataka koji se dobiju analizirajući podatke o stanovništvu (npr. dob, spol...).



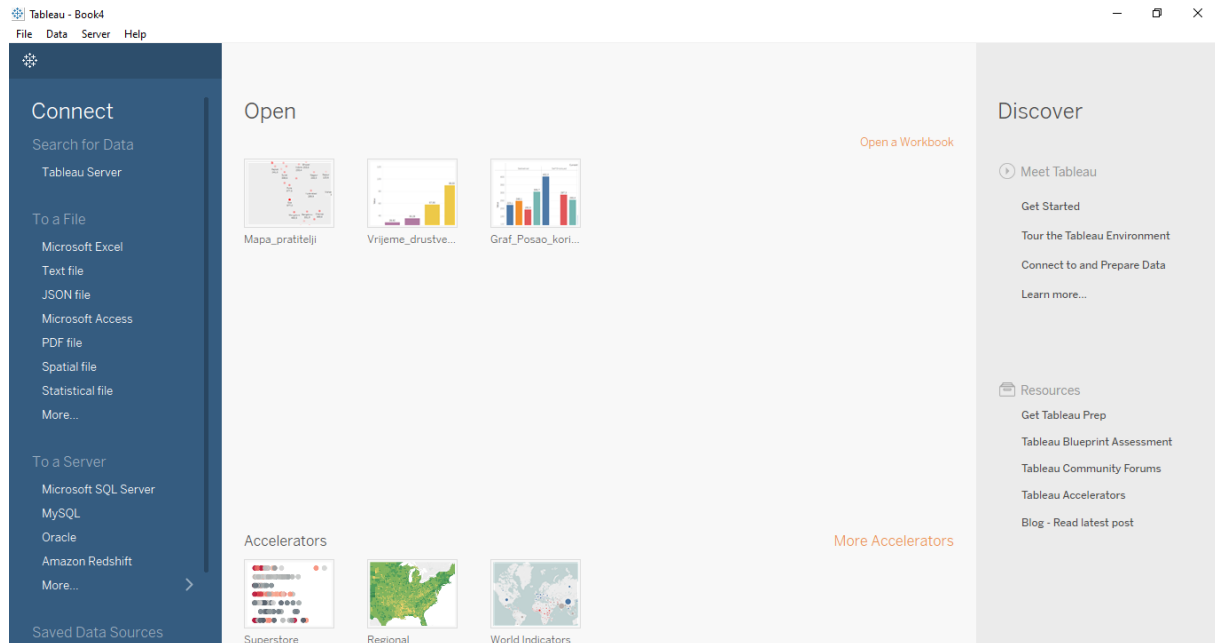
Slika 42 - Primjer histograma, preuzeto sa: (What is a Histogram Chart?, 2023)

Na ovom histogramu, x – os predstavlja ostvaren postotak na ispitu/testu, a y – os predstavlja broj studenata. Prema tome se može zaključiti kako je, na primjer, 8 studenata na nekom testu ostvarilo rezultat između 31%-40%.

3.4.1. Primjena vizualizacije u Tableau na vlastiti skup podataka

Sada kada su osnovni grafikoni objašnjeni u Tableau, obradit će se tablica iz 0. 3.3. Ulaganja u sport i korelacija tj. njene vizualizacije biti će prikazane u Tableau.

Prilikom otvaranja Tableau Desktop alata, njegovo sučelje prikazano je na slici 43:



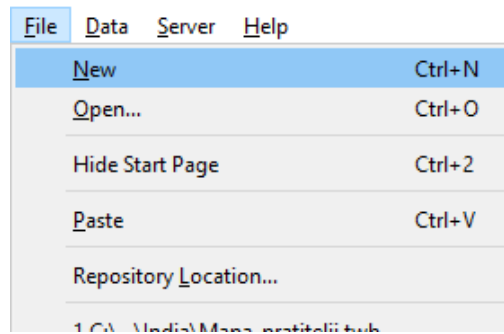
Slika 43 - Sučelje Tableau Desktop softvera

Kao što se vidi, neke su radne knjige već bile izrađene u ovom softveru (npr. „Mapa_pratitelji“) za izradu grafikona za neke prijašnje zadatke. Početna stranica sadži tri regije: *Connect*, *Open* i *Discover*. Pomoću *Connect* regije korisnik se može povezati s podacima sa spremljenih tablica ili datotekama drugog tipa (Microsoft Excel, tekstualne datoteke, JSON datoteka¹⁴ i sl.). Također, korisnik se može spojiti na bazu podataka i pristupiti podacima unutar nje (Microsoft SQL Server, Oracle..). Pod dijelom *Saved Data Sources* korisnik može vrlo lako pristupiti datotekama koje su spremljene u Tableau repozitorij. U *Open* dijelu, korisnik može otvoriti radne knjige s kojima je nedavno radio te prikvačiti (*engl. Pin*) radne knjige na početnu stranicu. Radne knjige koje su prikvačene na početnoj stranici pokazuju se na istoj i ako nisu nedavno korištene. *Discover* dio pruža korisniku mogućnost da jednostavno krene s radom u Tableau ako je početnik, da pronađe videa s kojima si može pomoći kako bi naučio raditi u softveru te čitanje najnovijih objava na Tableau blogu.

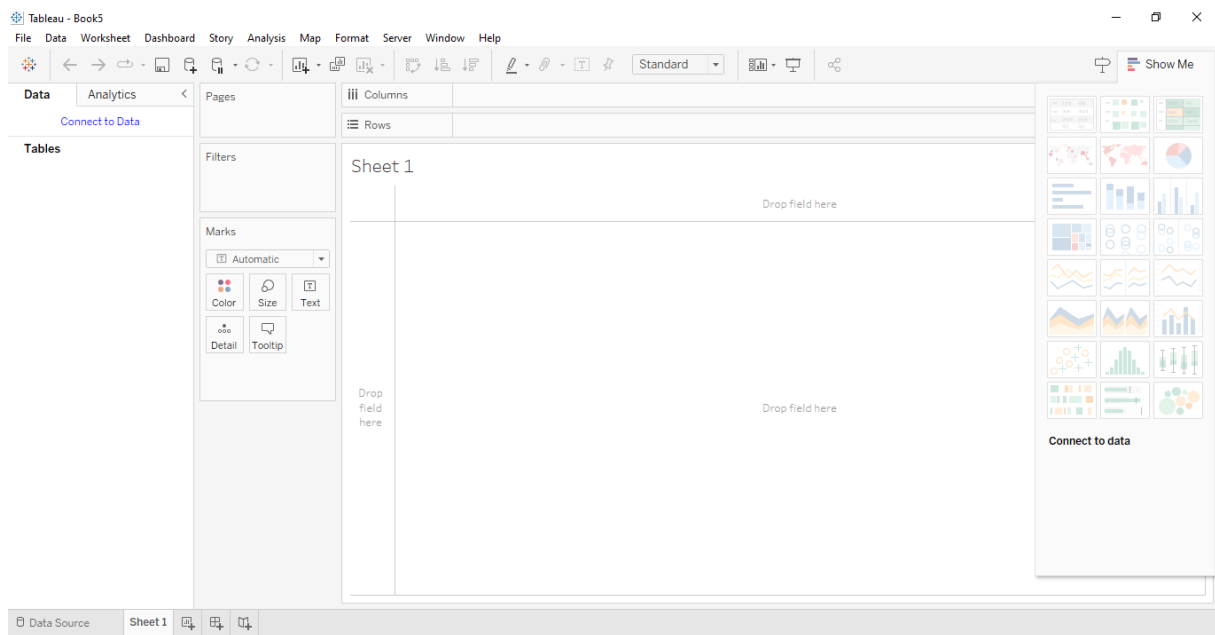
Da bi se kreirala nova radna knjiga, u desnom gornjem kutu klikne se *File* → *New* (slika 44). Zatim se otvori nova radna knjiga koja se po defaultu zove *Tableau – Book(n)*¹⁵. Sučelje nove radne knjige prikazano je na slici 45.

¹⁴ JavaScript Object Notation, format za pohranu i prijenos datoteka (W3Schools, 2023).

¹⁵ n – broj radne knjige, npr. Tableau – Book3

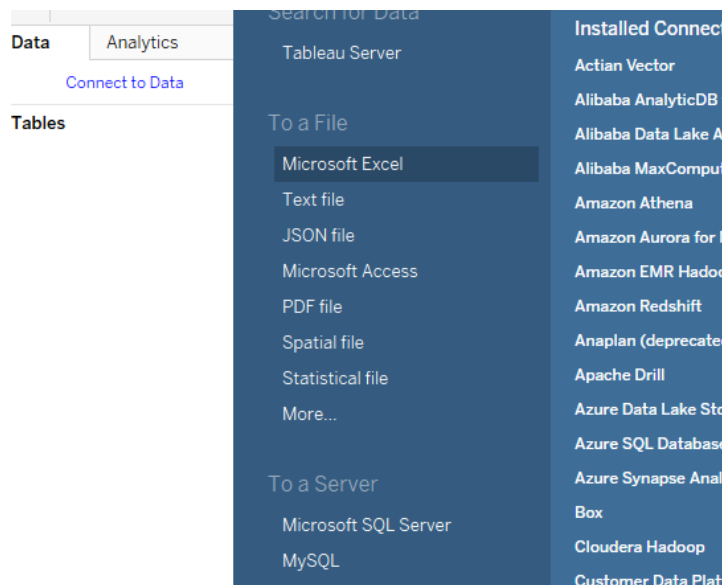


Slika 44 - Izrada nove radnje knjige

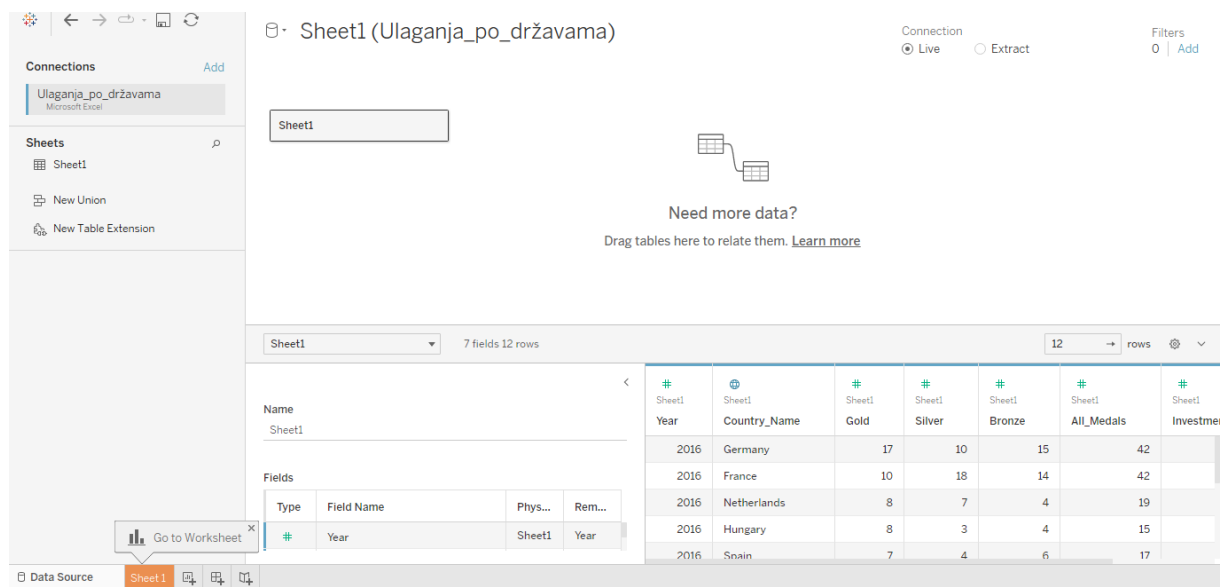


Slika 45 - Sučelje nove radne knjige

Da bi se podaci povezali sa softverom i kako bismo bili u mogućnosti izrađivati nove vizualizacije, kliknemo na *Connect to Data* te izabiremo s kojim se izvorom podataka želi povezati softver (slika 46). U ovom slučaju, to će biti *Microsoft Excel*.



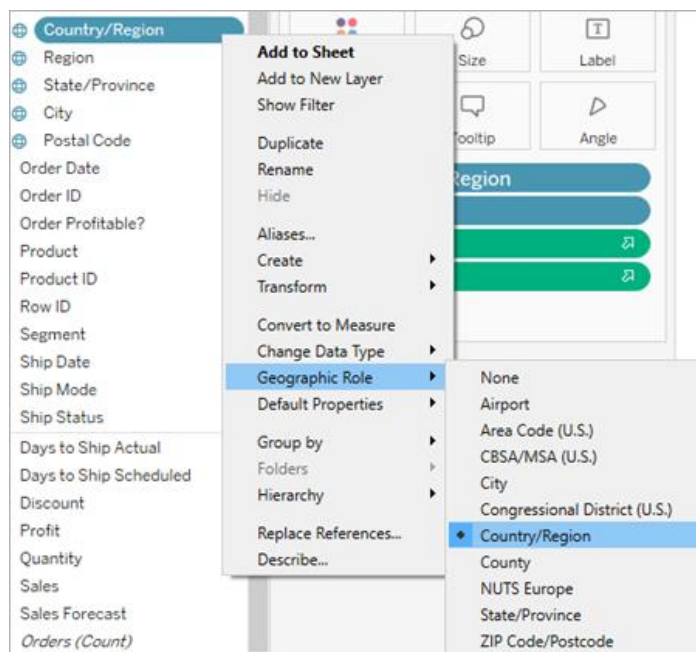
Slika 46 - Povezivanje podataka sa softverom



Slika 47 - Prikaz podataka u Tableau

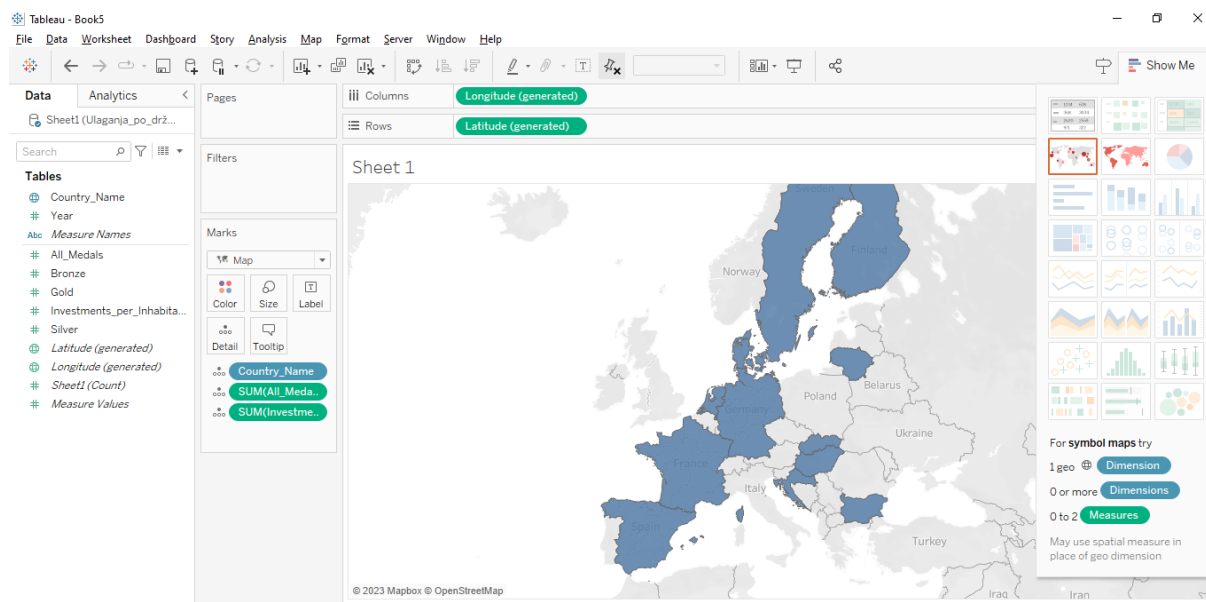
Za ovu vizualizaciju, nisu potrebne nikakve veze između podataka. Sljedeći korak je klik na „Go to Worksheet“ (slika 47) te odabir neke od vizualizacija koju želimo prikazati. Za to je potrebno podatke povući (*engl. Drag and Drop*) u odgovarajući prostor (*Columns* ili *Rows*). Vizualizacija koja je napravljena je karta Europe te su prikazane države koje se nalaze u originalnoj tablici. Da bi se dobila karta kao vizualizacija, u *Columns* i *Rows* bilo je potrebno povući automatski generiranu geografsku širinu i dužinu za pojedinu državu. Tableau u većini slučajeva svakom polju pridruži geografsku širinu

i dužinu. Ako se geografska širina i dužina automatski ne stvore, moguće ih je namjestiti u postavkama (slika 48).



Slika 48 - Mogućnost postavljanja geografske širine u dužine, preuzeto s: (A Guide To Charts, 2023)

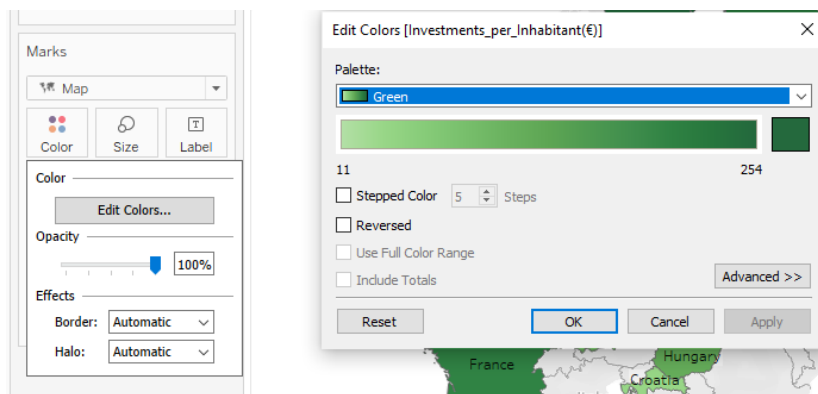
Pod karticom *Marks* (oznake) na slici 49 ispušteni su podaci o imenima država, ulaganjima po stanovniku te boja medalja koje su države osvojile.



Slika 49 - Podaci koje želimo prikazati

U kartici *Marks* postoji dio *Color* (boja) gdje se mijenja boja prikazane karte. Boja karte može biti jednobojna (sve zemlje predstavljene su istom bojom) ili sa prijelazima (gradijent). Da bi mogli države prikazati gradijentom, potrebno je pod *Marks* uhvatiti oznaku *SUM(Investments_per_Inhabitant(€))* te

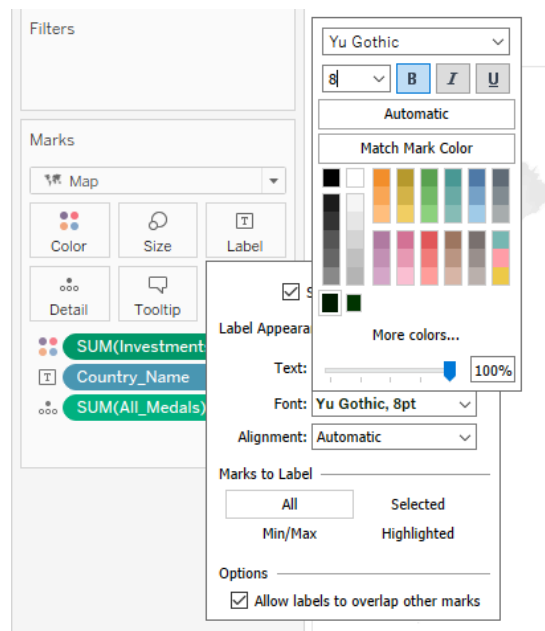
je ispustiti pod oznaku *Color*. Tada će biti moguće mijenjati gradijent i boju državama. Na slici 50 može se uočiti promjena boje na karti.



Slika 50 - Promjena boje za prikaz država

Boja je iz automatske plave promijenjena u zelenu. Moguće je i mijenjati i razinu gradijenta (*engl. Stepped Color*), tj. odrediti stupanj gradijenta. U primjeru se pustio automatski gradijent - 5 razina.

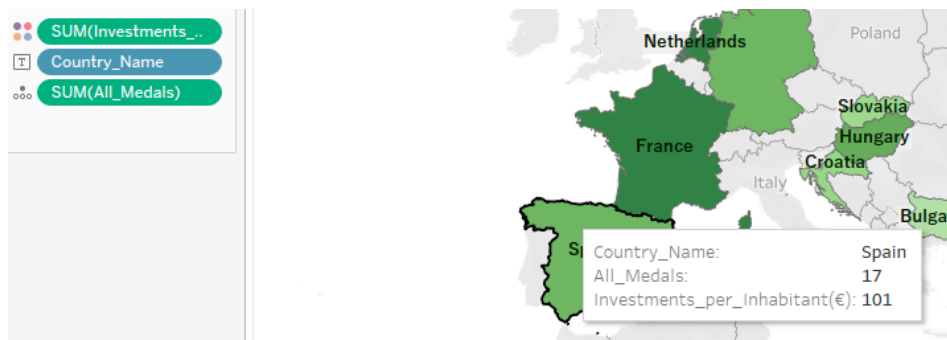
Za prikaz imena država na mapi, pod *Marks* je potrebno uhvatiti oznaku *County_Name* te ju ispustiti pod oznaku *Label*. Na slici 51 prikazana je mogućnosti mijenjanja boje, veličine i fonta teksta.



Slika 51 - Promjena fonta za prikaz imena država

Izabran je font *Yu Gothic*, veličine 8pt i podebljan (*engl. Bold*). Pod dijelom *Marks to Label* označeno je *All* kako bi sva imena država pisala na karti. Također, uključila se opcija „*Allow labels to overlap other marks*“ kako bi se prikazala sva imena država.

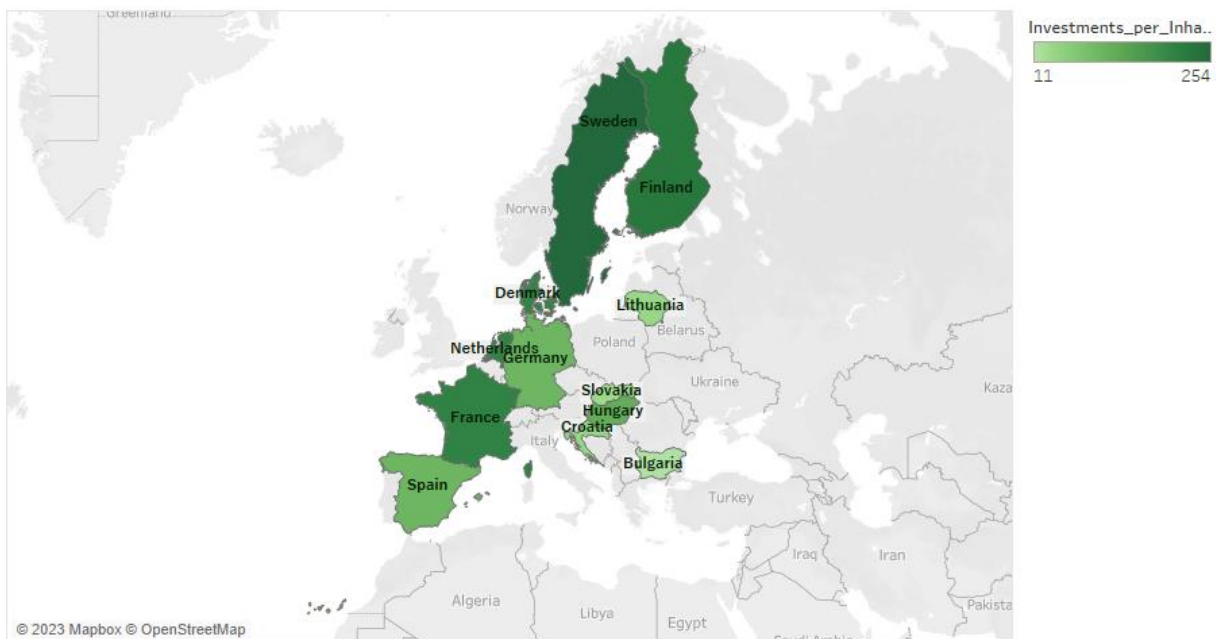
Pod *Detail* dijelom ispušten je stupac *SUM(All_Medals)* kako bi, kad se kursorom miša prijeđe preko određene zemlje, korisnik mogao vidjeti broj osvojenih medalja za pojedinu zemlju (slika 52).



Slika 52 - Prikaz detalja o svakoj državi

Na kraju je napravljenu vizualizaciju potrebno preimenovati (ako korisnik želi, preporučljivo zbog lakšeg snalaženja) i spremiti. Kada se na kraju rada, radna knjiga izveze u obliku slike (Worksheet→Export→Image), njena je vizualizacija prikazana na slici 53 te izgleda ovako:

Prikaz država



Slika 53 - Finalna vizualizacija Europskih zemalja i osvojenih medalja

4. Zaključak

U ovom radu opisane su različite metode analize i vizualizacije podataka koje pomažu u procesima istraživanja, analiziranja i razumijevanja raznolikih skupova podataka. Navedene metode nude dublji uvid u podatke, pomažu u predviđanju trendova te podržavaju donošenje informiranih odluka.

Odabir metode analize i vizualizacije ovisi o podacima s kojima korisnik radi. Bitno je izabrati metodu koja je podudarna s podacima tj. metodu koja na jasan i jednostavan način prikazuje rezultate analize podataka. Vizualizacija je ključni alat za prikazivanje i razumijevanje podataka. Grafikoni, tablice i druge vizualne tehnike pomažu u prepoznavanju obrazaca i prikazivanju rezultata drugima na jasan način.

Analiza i vizualizacija podataka ključni su za razumijevanje i iskorištavanje informacija iz velikih i složenih skupova podataka. Kombinacija različitih metoda omogućava donošenje bolje informiranih odluka, identifikaciju prilika i izazova te prikaz rezultata na jednostavan i razumljiv način.

Kod analize i vizualizacije podataka pripremila su se i koristila dva skupa: prvi skup podataka bio je tablica koja je sadržavala različite zemlje i broj osvojenih zlatnih, srebrnih te brončanih medalja, a u drugom skupu podataka su se nalazile neke Europske zemlje izvučene iz prve, originalne tablice, te je tom skupu podataka bio dodan još jedan stupac „Ulaganja po stanovniku“. Glavni cilj analize i vizualizacije podataka bio je analizirati podatke s različitih aspekata kako bi se ti podaci mogli vizualizirati na korisniku lako razumljiv način. Također, izračunata je i korelacija kako bi se ispitalo utječu li ulaganja po stanovniku na uspjeh pojedine države.

Iz provedene analize i vizualizacije podataka, daje se zaključiti da je korelacija uspjeha u sportu pojedine države u vidu broja osvojenih medalja na Olimpijskim igrama slabo povezana s ulaganjima te zemlje u sport po stanovniku (u Eurima). Izgleda da na uspjeh pojedine zemlje utječe mnogo više faktora, a ne samo količina uloženog novca u sport po stanovniku.

5. Popis slika

Slika 1 - Najpoznatija proširenja u VS Code – u, snimka zaslona napravljena sa: (Microsoft, Extensions, 2023)	9
Slika 2 - Korisničko sučelje VS Code-a, preuzeto sa: (Microsoft, User Interface, 2023).....	10
Slika 3 - Primjer vizualizacije u Tableau, vizualizaciju napravio: Tabelau Public korisnik Adam McCann, preuzeto sa: (TechTarget, 2019)	11
Slika 4 - NumPy array, preuzeto s: (Developers, 2022)	12
Slika 5 - Primjer histograma u Matplotlib-u, preuzeto sa: (Matplotlib, 2020)	13
Slika 6 - Preuzimanje prvog skupa podataka	14
Slika 7 - Prikaz tablice u Excelu	14
Slika 8 - Prikaz DataFrame - a, preuzeto s: (Getting Started, 2023)	15
Slika 9 - Dodavanje početnih biblioteka	16
Slika 10 - Učitavanje DataFrame-a	16
Slika 11 - Ispis nesortiranog DataFrame-a	16
Slika 12 - Funkcija sortiranja DataFrame-a.....	17
Slika 13 - Ispis sortiranog DataFrame-a	17
Slika 14 - Dodavanje novog stupca "All_Medals"	18
Slika 15 - Ispis novog stupca "All_Medals"	18
Slika 16 - Prikaz koda za pet država s najviše medalja	18
Slika 17 - Prikaz ispisa pet zemalja s najviše medalja	19
Slika 18 – Kod sa ispis zemlje s najviše medalja	19
Slika 19 - Ispis zemlje s najviše medalja.....	20
Slika 20 - Prikaz koda za pet država s najmanje medalja.....	20
Slika 21 - Prikaz ispisa pet zemalja s najmanje medalja	20
Slika 22 - Lista svih zemalja sa jednom osvojenom medaljom.....	20
Slika 23 - Ispis zemalja sa jednom medaljom	21
Slika 24 - Upit za zemlje koje nemaju osvojenu ni jednu medalju	21
Slika 25 - Ispis zemalja koje nemaju ni jednu medalju	21
Slika 26 - Kod za ispis rezultata Hrvatske.....	22
Slika 27 - Rezultati Hrvatske na ljetnim OI	22
Slika 28 – Grafikon napravljen pomoću Matplotlib – a	22
Slika 29 - Medalje za Hrvatsku kroz godine	23
Slika 30 - Tablica sa podacima o ulaganjima.....	24
Slika 31 - Dodavanje odgovarajućih biblioteka u novi Python File.....	25
Slika 32 - Učitavanje u DataFrame	26
Slika 33 - Izračunavanje i prikaz korelacije pomoću HeatMape.....	26
Slika 34 - Vizualizacija korelacije na Heatmapi	26
Slika 35 - Primjer tortnog grafikona, preuzeto s (Percentage pie chart for finance report, 2023).....	28
Slika 36 - Primjer stupičastog grafikona, preuzeto sa: (What is a Bar Chart?, 2022)	29
Slika 37 - Linijski grafikon s jednom linijom, preuzeto s: (Cisneros, 2020).....	30
Slika 38 - Linijski grafikon s više linija, preuzeto s: (Line Chart, 2023)	30
Slika 39 - Korelacija ne postoji na dijagramu raspršenosti, preuzeto s: (A Guide To Charts, 2023)....	31
Slika 40 - Savršena pozitivna korelacija na dijagramu raspršenosti, preuzeto s: (A Guide To Charts, 2023).....	31
Slika 41 - Savršena negativna korelacija na dijagramu raspršenosti, preuzeto s: (A Guide To Charts, 2023).....	32
Slika 42 - Primjer histograma, preuzeto sa: (What is a Histogram Chart?, 2023)	32
Slika 43 - Sučelje Tableau Desktop softvera.....	33
Slika 44 - Izrada nove radnje knjige.....	34

Slika 45 - Sučelje nove radne knjige	34
Slika 46 - Povezivanje podataka sa softverom	35
Slika 47 - Prikaz podataka u Tableau	35
Slika 48 - Mogućnost postavljanja geografske širine u dužine, preuzeto s: (A Guide To Charts, 2023)	36
Slika 49 - Podaci koje želimo prikazati	36
Slika 50 - Promjena boje za prikaz država	37
Slika 51 - Promjena fonta za prikaz imena država	37
Slika 52 - Prikaz detalja o svakoj državi	38
Slika 53 - Finalna vizualizacija Europskih zemalja i osvojenih medalja	38

6. Popis tablica

Tablica 1 - Usporedba različitih vrsta Tableau-a, tablica napravljena na temelju: (Simplilearn, 2023)	11
Tablica 2 - Kategorizacija snage korelacije u odnosu na korelacijski koeficijent, tablica napravljena na temelju: (Rahman, Ozair, Khan, & Sonkar, 2023, str. 184)	25

7. Literatura

- 125 years of Summer Olympics (Analysis & Visual)*. (2021). Dohvaćeno iz Kaggle:
<https://www.kaggle.com/code/hamdallak/125-years-of-summer-olympics-analysis-visual/input>
- A Guide To Charts*. (2023). Dohvaćeno iz Tableau: <https://www.tableau.com/data-insights/reference-library/visual-analytics/charts>
- Baesens, B., Bapna, R., Marsden, J., Vanthienen, J., & Zhao, J. (2016). Transformational Issues of Big Data and Analytics in Networked Business. *MIS Quarterly*, 807-818.
- Business Intelligence*. (August 2023). Dohvaćeno iz Wikipedia:
https://en.wikipedia.org/wiki/Business_intelligence
- Cisneros, M. (24. March 2020). Dohvaćeno iz StoryTelling Data:
<https://www.storytellingwithdata.com/blog/2020/3/24/what-is-a-line-graph>
- Cloud Storage*. (August 2023). Dohvaćeno iz Wikipedia: https://en.wikipedia.org/wiki/Cloud_storage
- Developers, N. (2022). *NumPy*. Dohvaćeno iz
https://numpy.org/doc/stable/user/absolute_beginners.html
- Getting Started*. (August 2023). Dohvaćeno iz Pandas:
https://pandas.pydata.org/docs/getting_started/index.html
- Grus, J. (2019). *Data Science from Scratch, First Principles with Python* (Svez. 2). (M. Cronin, Ur.) Sebastopol, Sjedinjene Američke Države: O'Reilly Media. Preuzeto 07 2023 iz
https://covid19.uthm.edu.my/wp-content/uploads/2020/04/Data-Science-from-Scratch-First-Principles-with-Python-by-Joel-Grus-z-lib.org_.epub_.pdf
- How much do governments spend on recreation and sport?* (23. September 2020). Dohvaćeno iz Eurostat: <https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/edn-20200923-1>

- Kaggle*. (29. August 2023). Dohvaćeno iz Wikipedia: <https://en.wikipedia.org/wiki/Kaggle>
- Korelacija*. (June 2023). Dohvaćeno iz Wikipedia: <https://hr.wikipedia.org/wiki/Korelacija>
- Kos Kavran, A., Kralj, A., & Ratković, M. (Rujan 2020.). *Sportski Marketing - 2.izdanje,skripta*. Preuzeto 07. 2023. iz Međimursko veleučilište u Čakovcu: <https://www.mev.hr/wp-content/uploads/2020/09/Sportski-marketing-2.-izdanje-skripta.pdf>
- Line Chart*. (2023). Dohvaćeno iz Excel Easy: <https://www.excel-easy.com/examples/line-chart.html>
- Matplotlib*. (June 2020). Dohvaćeno iz Python Charts: <https://www.pythoncharts.com/matplotlib/histograms/>
- Microsoft. (2023). *Extensions*. Preuzeto 30. 08 2023 iz Visual Studio Code: <https://marketplace.visualstudio.com/VSCode>
- Microsoft. (3. August 2023). User Interface. Preuzeto 31. 08 2023
- Percentage pie chart for finance report*. (2023). Dohvaćeno iz VectorStock: <https://www.vectorstock.com/royalty-free-vector/percentage-pie-chart-for-finance-report-template-vector-36321407>
- Rahman, F., Ozair, A., Khan, S., & Sonkar, A. (2023). Correlation. U F. Rahman, A. Ozair, S. R. Khan, & A. A. Sonkar, *Translational Surgery* (str. 183-189). Academic Press (Elsevier).
- Seaborn*. (2022). Preuzeto 13. September 2023 iz Seaborn: <https://seaborn.pydata.org/>
- Simplilearn*. (June 2023). Dohvaćeno iz What is Tableau: Power-packed Tutorial For Beginners: <https://www.simplilearn.com/tutorials/tableau-tutorial/what-is-tableau>
- Tableau Software*. (August 2023). Preuzeto 1. 9 2023 iz Wikipedia: https://en.wikipedia.org/wiki/Tableau_Software
- TechTarget. (27. June 2019). *Music data visualization*. Dohvaćeno iz TechTarget: <https://www.techtarget.com/searchbusinessanalytics/news/252465871/Music-data-visualization-comes-to-life-in-Tableau-collection>
- The MIT License*. (n.d.). Preuzeto 30. 08 2023 iz Open Source Initiative: <https://opensource.org/license/mit/>
- Visual Studio Code*. (August 2023). Dohvaćeno iz Wikipedia: https://en.wikipedia.org/wiki/Visual_Studio_Code
- W3Schools. (2023). Dohvaćeno iz W3Schools: https://www.w3schools.com/whatis/whatis_json.asp#:~:text=JSON%20stands%20for%20Java%20Script%20Object,server%20to%20a%20web%20page
- What is a Bar Chart?* (2022). Dohvaćeno iz Twinkl: <https://www.twinkl.com.au/teaching-wiki/bar-chart>
- What is a Histogram Chart?* (2023). Dohvaćeno iz Tibco: <https://www.tibco.com/reference-center/what-is-a-histogram-chart>