

SREDNJA ELEKTROTEHNIČKA ŠKOLA SARAJEVO



MATURSKI RAD

TEMA: Aplikacija za analizu podataka

Mentor :

Adnan Delić dipl. el. ing.

Učenik:

Ensar Hodžić

Sarajevo, april 2022.

Sadržaj

1. Uvod.....	3
1.1 Šta su podaci?.....	3
1.2 Podaci u računarstvu	3
2. Programski jezici.....	4
2.1 Šta su programski jezici?	4
2.2 C#.....	4
2.3 Python	4
2.3.1 Python paketi.....	5
2.3.2 Pandas	5
2.3.3. Matplotlib.....	5
2.3.4. Seaborn.....	5
2.4 Microsoft Visual Studio C# 2010	6
2.5 Microsoft Visual Studio Code	6
3. Podaci.....	7
3.1 Oblici podataka	7
3.2 Nauka o podacima.....	8
3.3 Deskriptivna statistika podataka	8
3.4 Grafički prikaz podataka.....	9
3.4.1 Stupičasti graf	9
3.4.2 Linijski graf.....	10
3.4.3 Kružni graf	10
3.4.4 Raspršeni graf	11
4. Programski kod	12
4.1 Uvod.....	12
4.2 Forma za odabir grafika	12
4.2 Forma za podešavanje grafika	13
4.3 Prikaz grafika	14
5. Zaključak.....	15
6. Mišljenje o radu	16
7. Literatura	17

1. Uvod

1.1 Šta su podaci?

Podaci općenito predstavljaju individualne činjenice, statistike ili stavke informacije, često numeričkog tipa. Kada bi htjeli dati neku formalniju definiciju, podaci predstavljaju skup vrijednosti nekih kvalitativnih ili kvantitativnih varijabli koje se vežu za neku osobu ili predmet.

1.2 Podaci u računarstvu

Podaci u računarstvu predstavljaju bilo koji niz jednog ili više simbola. Kako bi podaci postali neka informacija, potrebno ih je tumačiti. Digitalni podaci su podaci koji su predstavljeni koristeći binarni brojni sistem (0 i 1), te u današnjim modernim računarskim sistema, svi podaci su digitalni.

Podaci koji predstavljaju neke vrijednosti, karaktere ili simbole nad kojima kompjuter može izvršavati različite operacije se spašavaju na različitim tipovima računarske memorije, poput magnetnih ("hard" diskovi), optičkih (diskovi), elektroničkih ("solid state" diskovi) i drugih oblika medija. Podaci se unose u računar, kao i izvoze iz računara putem ulazno-izlaznih jedinica.

Sam koncept podataka u kontekstu računarstva potiče od Claude Shannon-ovog rada. Claude Shannon je američki matematičar koji je poznat kao "otac teorije o informacijama". On je uveo binarne digitalne koncepte koji se zasnivaju na primjeni "Boolean" logici u elektroničkim kolima. Boolean predstavlja logički tip podatka koji ima samo dvije vrijednosti, true (tačno) ili false (netačno), koji se često prikazuju brojčanim zapisom kao 1 i 0, respektivno.

Računari predstavljaju podatke, uključujući slike, zvukove, video ili tekst, kao binarne vrijednosti koristeći samo dva broja: 1 i 0. Bit je najmanja mjerna jedinica koja označava količinu podataka i predstavlja samo jednu vrijednost. Današnji računari imaju mogućnost spašavanja i čitanja gigabyte-ima (2^{30} bita), terabyte-ima (2^{40} bita), pa čak i petabyte-ima (2^{50} bita) podataka.

2. Programski jezici

2.1 Šta su programski jezici?

Programski jezik predstavlja jedan umjetni jezik, kreiran u svrhu da olakša korisnicima upravljanjem računara. Programski jezik se sastoji od određenih sintaksičkih skupina, simbola i pravila njihovog slaganja koja su jednostavna za shvatiti korisnicima. Programski jezik ima zadatak da prevede te sintaksičke skupine i simbole u skup upustava koje procesor može izvršiti. Osnovna podjela je na niže (strojne jezike) i više (orijentirane ljudima) programske jezike.

2.2 C#



C# je objektno orijentirani programski jezik razvijen od strane kompanije Microsoft, s ciljem da .NET platforma dobije jednostavan, siguran, moderan, objektno orijentiran programski jezik visokih performansi. C# je nastao kao unapređena verzija programskih jezika Java, C, i C++. Najveća prednost programiranja u C# u odnosu na jezike iz kojih je proistekao je znatno lakša upotreba. C# je idealan za sve one koji žele da savladaju vještine programiranja u .NET frameworku korištenjem Visual Studio razvojnog okruženja.

2.3 Python



Python je visoki programski jezik opće namjene. U današnje vrijeme je vrlo popularan zato što je veoma jednostavan za naučiti i koristiti. Za razliku od jezika poput C, C++ i ostalih, Python je interpretirani programski jezik, sličan Javi, i česta kritika jeste da su takvi jezici spori. Međutim unatoč toj mani, Python je veoma popularan u današnjoj industriji, najčešće kao back-end programski jezik. Njegove prednosti su to što ima jednostavnu strukturu sintakse, veliki broj dodatnih paketa, neovisnost o platformi i pogodan je za mašinsko učenje i vještačku inteligenciju.

2.3.1 Python paketi

Python paketi predstavljaju kolekciju povezanih modula koji imaju za cilj da olakšaju programiranje. Paketi sadrže različite skupove koda koji se mogu koristiti u više različitih programa, što znači da ne moramo ponavljati iste linije koda u različitim projektima. Ugrađeni upravitelj ovim paketima u Pythonu se zove PIP, i pomoću njega možemo instalirati dodatne pakete. Prednost Pythona jeste što sadrži mnoštvo dodatnih paketa za bilo koju vrstu namjene.

2.3.2 Pandas



Pandas je Python paket koji je popularan za korištenje u nauci o podacima, analizi podataka, kao i zadataka što se tiču mašinskog učenja. Pandas je prvi put predstavljen 11. januara 2008. godine, njegov kreator se zove Wes McKinney, međutim pošto je pandas “open-source” paket, Python zajednica je konstantno poboljšavala ovaj paket sve do danas.

2.3.3. Matplotlib



Matplotlib je najpopularniji Python paket za vizualizaciju podataka i kreiranje različitih vrsta grafika. On je nastao kao grafička ekstenzija jednom drugom popularnom paketu NumPy. Matplotlib je dizajniran tako da kreiranje grafika bude što jednostavnije i da je potrebno što manji broj linija koda kako bi se to ostvarilo.

2.3.4. Seaborn



Seaborn je Python paket koji se nadograđuje na Matplotlib i također služi za kreiranje različitih vrsta grafika. Njegova prednost na odnosu Matplotlib-a jeste da je dizajniran da radi sa pandas strukturama podataka. Kako bismo koristili Seaborn u našim projektima, morami imati i Matplotlib paket instaliran.

2.4 Microsoft Visual Studio C# 2010



U oktobru 2009. godine izašao je Windows 7, jedan od najpopularnijih operativnih sistema u historiji. 12. aprila 2010. godine, Microsoft je predstavio novu verziju popularnog integrisanog razvojnog okruženja Visual Studio 2010. Ova nova verzija je bazirana na .NET frameworku verzije 4.0, koja donosi podršku za novije računarske sisteme sa više jezgrenim procesorima. Visual Studio se koristi za razvoj Windows programa, web stranica, web aplikacija kao i mobilnih aplikacija. Popularno je razvojno okruženje zato što sadrži podršku za mnoge različite programske jezike, jednostavan interfejs za pravljenje grafičkih interfejsa aplikacija kao i urednik koda sa ugrađenom IntelliSense funkcijom. IntelliSense je Microsoftova implementacija automatskog predlaganja koda i veoma je koristan alat koji pojednostavljuje i ubrzava pisanje koda u različitim programskim jezicima.

2.5 Microsoft Visual Studio Code



Microsoft Visual Studio Code je drugačiji program od standardnog Visual Studio. Visual Studio Code (ili VS Code) je besplatni uređivač koda koji programeru pomaže u pisanju koda, u uklanjanju grešaka i popravljanju koda koristeći IntelliSense tehnologiju. Razlika između Visual Studio i VS Code jeste da je Visual Studio čitavo integrisano razvojno okruženje koje sadrži mnoge različite alate, dok VS Code predstavlja samo uređivač koda. Veoma je popularan zato što je to besplatan program koji ima mnoge prednosti poput podrške za više različitih programskih jezika, korištenje IntelliSense tehnologije, podrške među različitim operativnim sistemima kao i podršku za različita proširenja.

3. Podaci

3.1 Oblici podataka

Kada se bavimo sa velikim količinama podataka, važno nam je da znamo sa kojom vrstom podataka se bavimo, kako bi uradili tačnu analizu podataka kojim se bavimo da bi dobili što bolje i tačnije rezultate. Postoje dvije glavne podjele vrste podatka: kvalitativni i kvantitativni podaci.

Kvalitativni ili kategorični podaci su podaci koji opisuju neki predmet. To znači da ova vrsta podataka se ne može brojati ili mjeriti koristeći brojeve i zato se dijele u različite kategorije. Primjer ove vrste podataka je kada želimo opisati spol neke osobe. Spol može biti “muško”, “žensko” ili “ostalo”. Kvalitativni podaci se također dijele na nazivne i redne kvalitativne podatke. Nazivni podaci su podaci koji nemaju neki prirodni poredak, naprimjer kada imamo podatak o boji nečega. Ne možemo porediti jednu boju sa drugom ili ih na neki način poredati. Dok redni podaci imaju neki prirodni poredak, poput nekog reda veličina, od najmanjeg ka najvećem.

Kvantitativni podaci su podaci koji pokušavaju kvantificirati predmete i to rade koristeći numeričke vrijednosti što ih po prirodi čine brojivim. Primjeri su cijena nekog telefona, vrijednost nekog popusta, broj recenzija nekog artikla, itd. Kao i kvalitativni podaci, imamo i dvije vrste kvantitativnih podataka, diskretne i kontinualne podatke. Diskretni podaci predstavljaju brojne vrijednosti koje su cijeli broj, to jeste nemaju decimalnih mjesta. Koristimo ih često kada opisujemo neke stalne vrijednosti. Primjer ovih podataka je broj zvučnika na telefonu, jezgri procesora, itd. Kontinualni podaci predstavljaju brojne vrijednosti koje pripadaju skupu realnih brojeva. To znači da mogu imati decimalne vrijednosti. Koristimo ih često kada opisujemo neku vrijednost koja se stalno mjenja. Primjer ovih podataka je radna frekvencija nekog procesora, promjena dnevne temperature, itd.

3.2 Nauka o podacima

Nauka o podacima predstavlja jedno znanstveno područje koje primjenjuje znanstvene metode, procese, algoritme i sisteme kako bi izvukli neko znanje i uvide iz različitih vrsta podataka i da primjene to znanje dobiveno iz tih podataka na široko područje primjene. Ona predstavlja jedan koncept koji ujedinjuje statistiku, analizu podataka, informatiku i njihove metode kako bi razumjeli i analizirali pojave u podacima. Koristi različite tehnike i teorije iz mnogih različitih nauka poput matematike, statistike, računarske nauke, itd. Nauka o podacima je usko povezana sa drugim naukama, poput rudarenja podacima, mašinskog učenja i “velikim” podacima. Naučnik o podacima je osoba koja piše kod i kreira programe u kojima primjenjuje znanje iz statistike kako bi napravio uvide iz podataka.

3.3 Deskriptivna statistika podataka

Deskriptivna statistika je dio nauke analize podataka koja ima zadatak da opiše, pokaže ili sumira podatke na neki značajan način. Sastoji se od kalkulacija različitih mjera poput mjerenja centra, mjerenja promjene, percentila i građe tabela i grafikona.

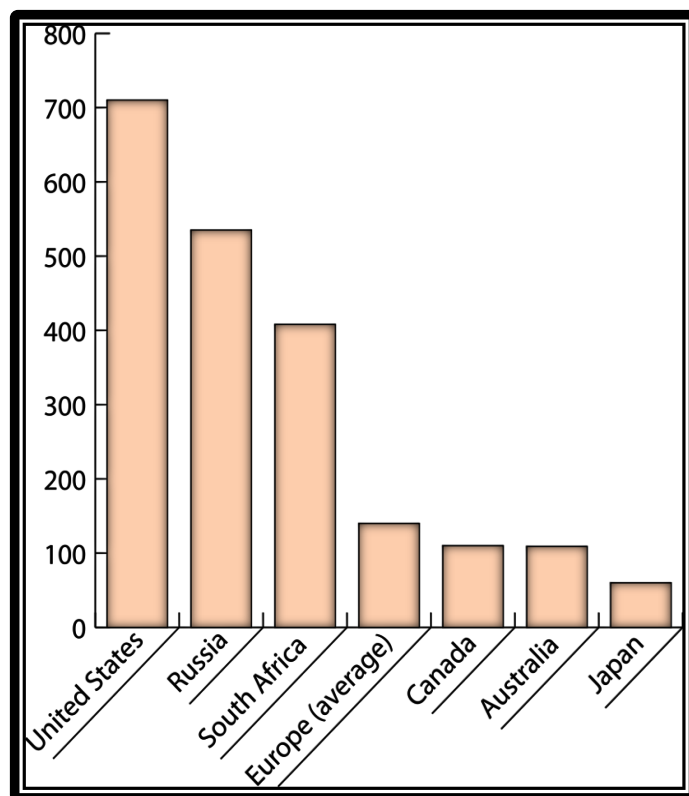
Za deskriptivnu statistiku podataka imamo različite mjere koje računamo kako bi opisali podatke. Aritmetička sredina podataka podrazumijeva dijeljenje sume svih kvantitativnih vrijednosti sa brojem vrijednosti. Medija je vrijednost koja određuje sredinu distribucije vrijednosti podataka. Za razliku od aritmetičke sredine, medijan je manje ostjetljiv na ekstremne vrijednosti, te je čini posebno pogodnom za nepravilne asimetrične distribucije.

3.4 Grafički prikaz podataka

Često su podaci pregledaniji ako su predstavljeni grafički. Danas imamo mnoge programe koje posao grafičkog predstavljanja podataka rade za nas. Postoje brojne grafičke mogućnosti kada je riječ o prikazivanju podataka. Ja ću ovdje objasniti samo četiri.

3.4.1 Stupičasti graf

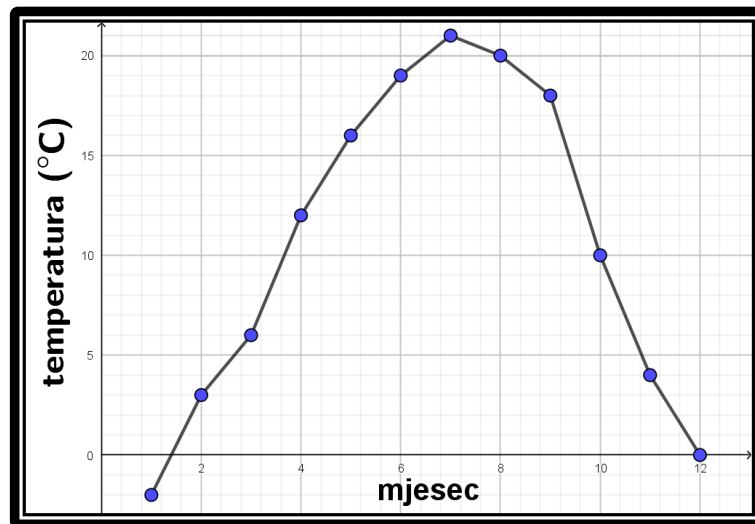
Stupičasti graf je graf koji se sastoji od pravouglanih oblika, čije su dužine proporcionalne vrijednostima koje oni predstavljaju. Koriste se za uspoređivanje dvije ili više vrijednosti. Pomoću stupičastih grafova možemo dobiti vizualni prikaz usporedbe skupina podataka ili učestalosti pojavljivanja različitih osobina podataka.



Slika 1: Primjer stupičastog grafa

3.4.2 Linijski graf

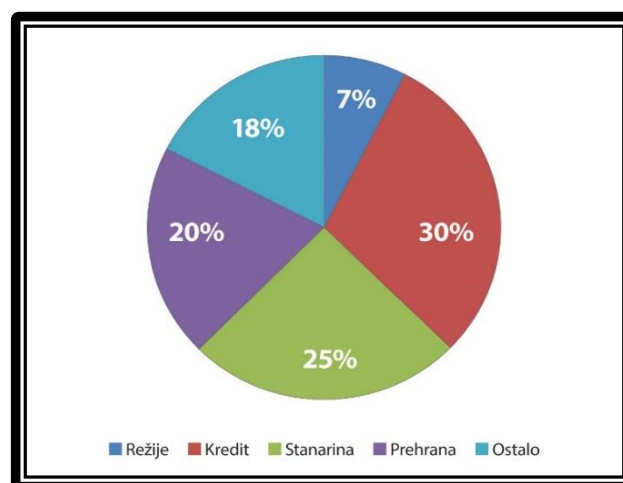
Linijski grafikoni se koriste za prikazivanje međuodnosa dviju promjenljivih varijabli tijekom vremena. Primjeri veličina koje se prikazuju linijskim grafikonom su promjena temperature, količine padavina, cijena i slično.



Slika 2: Primjer linijskog grafa

3.4.3 Kružni graf

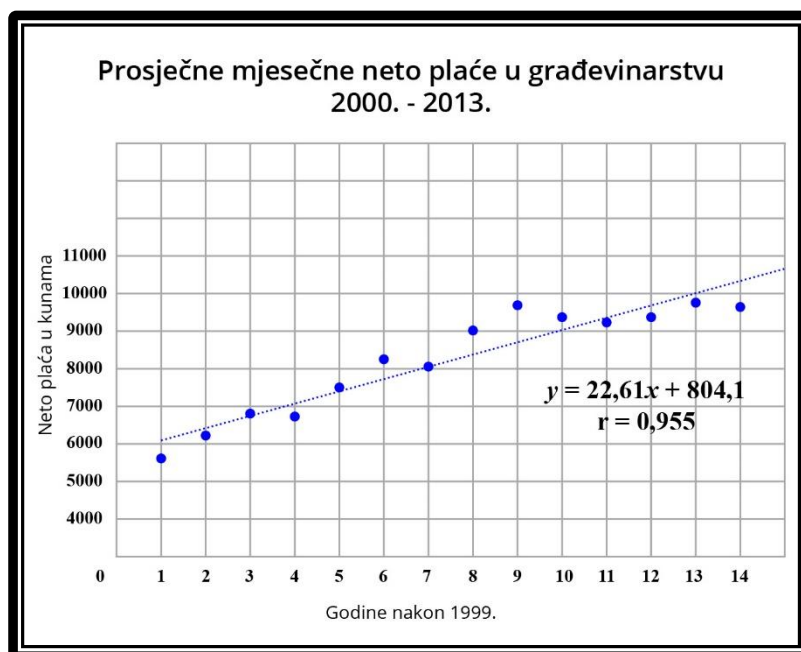
Kružni graf je način grafičkog prikazivanja podataka u kojem su vrijednosti podataka prikazane veličinama kružnih isječaka. Veličina kružnih isječaka je proporcionalna vrijednostima koje predstavljaju. Kružnim grafom prikazuju se dijelovi cjeline ili odnosi dijelova prema cjelini, a najčešće se izražavaju u postotcima.



Slika 3: Primjer kružnog grafa

3.4.4 Raspršeni graf

Raspršeni graf prikazuje odnos dvije varijable, u obliku X i Y koordinata, koji predstavlja tačku u koordinatnom sistemu. Najčešće se koristi za prikaz dvije varijable u velikom skupu podataka. Raspršeni grafikoni obično se koriste za ilustraciju i usporedbu brojčanih vrijednosti, kao što su znanstveni, statistički ili tehnički podaci.



Slika 4: Primjer raspršenog grafa

4. Programski kod

4.1 Uvod

Moja aplikacija koristi dva programska jezika, C# i Python. C# se koristi kako bi se napravio elegantan i jednostavan grafički interfejs za odabir i podešavanje grafika. Python se koristi sa samu kreaciju grafika, uz pomoć dva dodatna Python paketa, Matplotlib i Seaborn.

4.2 Forma za odabir grafika

The screenshot shows a window titled 'Analizator podataka' with a light blue background. It contains three main sections for user input:

- 1. Izaberite vrstu grafika**: This section offers four options, each with a small icon and a description:
 - Linijski graf**: Represented by a line graph icon. Description: 'Ovi grafikoni se koriste kada imate podatke koje se mijenjaju konstantno tokom vremena. Najčešće poredimo neke vrijednosti u određenom vremenu.'
 - Stupčasti graf**: Represented by a bar chart icon. Description: 'Ovi grafikoni se koriste kada imate kategorijske podatke. Jedna osa grafika predstavlja kategoriju mjerenih podataka, druga osa predstavlja vrijednost.'
 - Tortni graf**: Represented by a pie chart icon. Description: 'Ovi grafikoni se koriste kada želite prikazati sastav nečega. Potrebno je izabrati listu kategorija i numeričkih vrijednosti.'
 - Raspršeni graf**: Represented by a scatter plot icon. Description: 'Ovi grafikoni se koriste kada želimo prikazati poređenje dva para vrijednosti koje su povezane jedna sa drugom.'
- 2. Izaberite vaše podatke (.csv)**: This section includes a text box containing the file path 'C:\Users\Ensar\Downloads\Netflix Stocks Capstone\Netflix Sto' and an 'Otvori' button.
- 3. Izaberite kolone koje želite uporediti (maksimalno dvije)**: This section features a list of column names with checkboxes: 'Date', 'Open', 'High', 'Low', 'Close', 'Adj Close', and 'Volume'. To the right of this list, there is an instruction: 'Za ovaj grafikon, prvo izaberite vrijednosti za x-osu (najčešće vrijeme), a drugo izaberite vrijednosti koje se mijenjaju'.

At the bottom center of the window is a 'Nastavite dalje' button.

Slika 5: Forma za odabir grafika

U ovoj formi prvi imamo izbor vrste grafika koji želimo kreirati. Nakon odabira vrste grafika, birmo podatke koje želimo koristiti. Nakon odabira podataka, prikazuje se prikaz naziva kolona unutar naše tabele, gdje birmo dvije kolone za stvaranje grafika. Pored prikaza naziva, nalazi se dodatno uputstvo kako izabrati vrijednosti.

Kako bi prikazao nazive kolona u podacima, koristim Python. Kada izaberete vaše podatke, C# spašava putanju tog fajla u jedan tekstualni fajl. Onda pokreće Python skriptu, koja iz tog tekstualnog fajla čita putanju vašeg fajla sa podacima, otvara vaše podatke, čita nazive kolona i spašava ih u jedan drugi tekstualni fajl, koji C# opet otvara i prikaže na vašem ekranu.

```
import pandas as pd

file=open("path.txt", "r")
path = file.read()
file.close()
df = pd.read_csv(path[:len(path)-1])
newfile = open("names.txt","w")
for col in df.columns:
    newfile.write(col+",")
newfile.close()
```

Slika 6: Python kod za čitanje naziva kolona

4.2 Forma za podešavanje grafika

Analizator podataka

Podesite izgled vašeg grafika

Izaberite boju teksta
Izaberi

Izaberite paletu boja
▼

Napišite naslov x-ose

Napišite naslov grafa

Izaberite boju pozadine
Izaberi

Izaberite boju pozadine grafika
Izaberi

Napišite naslov y-ose

Odredite rotaciju oznaka
0

Kada pritisnete "Kreiraj", otvoriće vam se prozor sa vašim grafikom.
Povećavanjem prozora podešavate veličinu grafika
U tom prozoru imate dugme gdje možete spasiti grafik kao sliku.
Ukoliko želite dodatno podešavati grafik samo zatvorite prozor.

Nazad Kreiraj

Slika 7: Forma za podešavanje postavki grafika

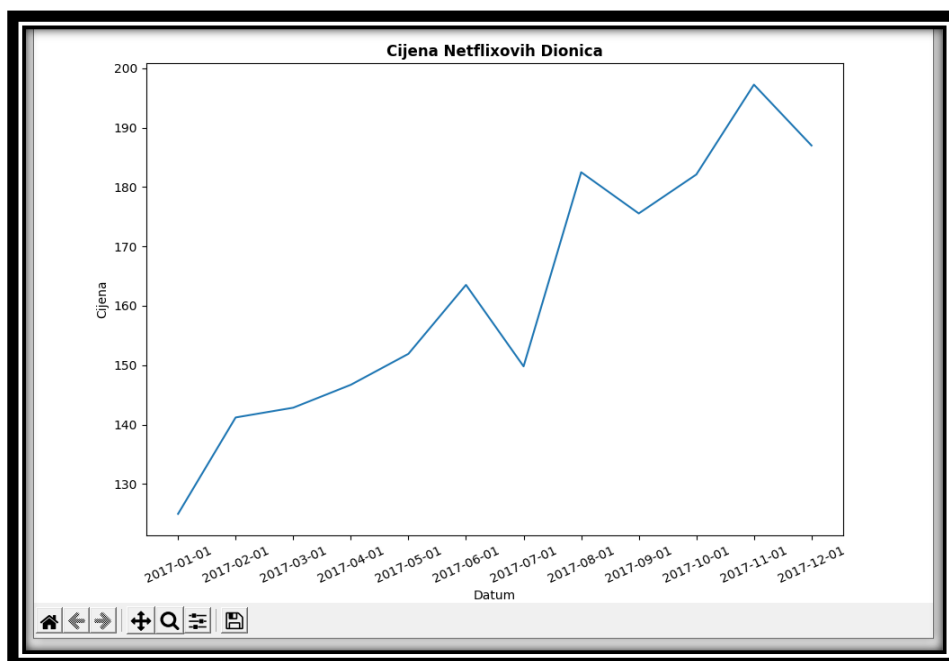
U ovoj formi se nalaze sve opcije sa kojima možemo podešavati kako će završni grafik izgledati. Postavke poput boje pozadine grafika, naslov grafika, rotacije oznaka, itd. Kao i u prošloj formi, C# sve ove postavke spašava u jedan tekstualni fajl koji će Python skripta čitati. Zaviseći od odabrane vrste grafika, pokreće se određena skripta. Primjer koda skripte za kreiranje stupičastog grafika u Pythonu koristeći Seaborn se nalazi ispod.

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

data = pd.read_csv("sales.csv")
sns.barplot(x="Date", y="Price", data=data)
plt.xlabel("Date of the sales")
plt.ylabel("Prices")
plt.title("Yearly sales")
plt.show()
```

Slika 8: Primjer koda za kreiranje stupičastog grafika

4.3 Prikaz grafika



Slika 9: Primjer napravljenog grafika

Kada pritisnete dugme Kreiraj, otvoriti će se drugi prozor sa vašim grafikom, gdje možete mjenjati veličinu prozora da mjenjate veličinu grafika. Također imate dugme za spašavanje grafika kao sliku.

5. Zaključak

U današnjem vremenu, podaci su ključan faktor. Kompanije skupljaju što više podataka o njihovim korisnicima kako bi prilagodili njihove proizvode, prikazivali reklame ciljane na određene skupine, ili kako bi predvidjeli šta im je potrebno u datom trenutku.

Nauka o podacima je veoma interesantna nauka trenutno, jer buduće tehnologije zavise o podacima. Koncepti poput mašinskog učenja ili vještačke inteligencije se zasnivaju na velikim količinama sortiranih podataka iz kojih računari mogu razumjeti i praviti odluke na osnovu njih.

Vizualni prikaz podataka služi kao jedan ključan faktor za analizu podataka, jer ljudski mozak ima sposobnost da vizuelno primjeti neke šablone ili povezanost među podacima ukoliko su prikazani vizuelno. Također je jednostavnije razumjeti date podatke običnim ljudima ako su predstavljeni grafički.

6. Mišljenje o radu

Potpis mentora:

Adnan Delić dipl. el. ing.

7. Literatura

- <https://www.intellspot.com/types-graphs-charts/>
- <https://datavizcatalogue.com/>
- <https://www.techtarget.com/searchdatamanagement/definition/data>
- https://en.wikipedia.org/wiki/Data_science
- <https://gocoding.org/hr/what-is-python-programming/>
- https://en.wikipedia.org/wiki/Visual_Studio_Code
- https://en.wikipedia.org/wiki/Microsoft_Visual_Studio
- <https://empiricom.ba/modul-9-c/>