

Student: Balša Dogandžić

Broj dosijea: 20/124i

Podgorica, septembar 2023. godine

Univerzitet Donja Gorica

Fakultet za informacione sisteme i tehnologije

Podgorica

**Prikaz rezultata analize podataka na veb-u pomoću Python programskog jezika sa konkretnim primjerom**

Diplomski rad



Mentor: mr Stevan Čakić

Student: Balša Dogandžić

Broj dosijea: 20/124i

Student: Balša Dogandžić

Broj dosijea: 20/124i

Podgorica, septembar 2023. godine

Univerzitet Donja Gorica

Fakultet za informacione sisteme i tehnologije

Podgorica

**Prikaz rezultata analize podataka na veb-u pomoću Python programskog jezika sa konkretnim primjerom**

Diplomski rad

# **APSTRAKT**

Python je jedan od najpopularnijih programskih jezika zbog svoje jednostavnosti, ali i zbog njegovih ogromnih mogućnosti. U ovom radu je opisan sistem koji je napravljen upravo pomoću ovog programskog jezika. Sistem predstavlja veb sajt napravljen pomoću Django paketa koji prikazuje rezultate analize i grafičke vizualizacije podataka realizovane pomoću paketa Pandas, Matplotlib i Seaborn na interfejsu pretraživača. Nakon opisa sistema u poglavlju diskusije će biti izneseni neki slučajevi korišćenja ovakvih sistema, ali će biti raspravljano i tome da li je ovaj sistem dobar i na koji način može da se poboljša.

**Ključne riječi**: Python, Web, Podaci, Analiza, Vizualizacija, Django, Pandas.

# **ABSTRACT**

Python is one of the most popular programming language because of it’s simplicity, but also because of it’s great versatility. A system described in this thesis is made with previously mentioned programming language. System is a web application made with Django framework which displays data analysis results and graphic visualizations of data with Pandas, Matplotlib and Seaborn packages on browser’s interface. After system description, in the discussion chapter some use cases of the similar systems will be presented, but it will be discussed whether this system is good and how to improve it.

**Key words**: Python, Web, Data, Analysis, Visualization, Django, Pandas.

# SADRŽAJ

[APSTRAKT 2](#_Toc143288108)

[ABSTRACT 2](#_Toc143288109)

[SADRŽAJ 3](#_Toc143288110)

[Lista slika 3](#_Toc143288111)

[1. UVOD 4](#_Toc143288112)

[1.1 Ideja rada i cilj rada 5](#_Toc143288113)

[1.2 Očekivanja od rada 5](#_Toc143288114)

[1.3 Tema u okviru mreže međuzavisnosti 5](#_Toc143288115)

[2. Metodologija 5](#_Toc143288116)

[2.1 Softverski paketi 6](#_Toc143288117)

[LITERATURA 7](#_Toc143288118)

# **Lista slika**

**No table of figures entries found.**

# **1. UVOD**

Podaci su svuda oko nas i oni su osnova svih sistema koji nam olakšavaju svakodnevni život. Oni nam omogućavaju da u njima vidimo neke pojave, identifikujemo potencijalne problema, kao i da donesemo odgovarajuće odluke u biznisu ili drugim sferama života. Podataka je iz godine u godinu sve više, prema istraživačima iz CISCO organizacije: protok podataka kroz internet 2022 godine se procjenjuje na 4.8 zetabajta, što je oko 4.8 \* 1021 bajtova. (Barnett, T.; Jain, S. (2018). Cisco visual networking index (vni) complete forecast update, 2017–2022. Americas/EMEAR Cisco Knowledge Network (CKN) Presentation, strana br. 8) Rastom protoka podataka raste i potreba da se ovi podaci analiziraju, i da se iz njih stvori neka nova vrijednost. Osim što je tehnologija u velikoj mjeri i “krivac” za generisanje ovolike količine podataka, ona predstavlja i rješenje kako da se ovi podaci predstave na razumljiv način.

Postoji veliki broj softverskih rješenja bilo to komercijalnih ili besplatnih rješenja otvorenog koda, neki od njih su: MS Excel, R programski jezik, Matlab, Scala, Python i mnogi drugi. Python i MS Excel su svakako dva najkorišćenija i najpoznatija alata za obradu i manipulaciju nad podacima. Prednost Python-a u odnosu na Excel je ta što je Python programski jezik otvorenog koda i kao programski jezik šire namjene nije odgraničen samo na rad sa podacima. Sa Python programskim jezikom je moguće kreirati veb aplikacije (Django, Flask), desktop aplikacije (Tkinter), ali i skripte različitih namjena pomoću ogromnog broja paketa. Nedostatak Python-a u odnosu na MS Excel i ostale komercijalne softvere je taj što za korišćenje Python-a korisnik mora posjedovati programersko znanje, dok komercijalni alati korisniku pružaju grafički interfejs koji mu omogućava lakše korišćenje softvera i bolje korisničko iskustvo. Ali i pored tih nedostataka Python sa svojim paketima za analizu podataka (NumPy, Pandas, Matplotlib...) dobija sve veću popularnost zbog svoje jednostavnosti, brzine i potencijala. Između ostalog je i to razlog zašto je upravo ovaj programski jezik tema ovog rada. U narednim poglavljima ovog rada će biti opisan praktični dio projekta za čiju realizaciju su korišćeni Python paketi za analizu podataka koji su prethodno pomenuti, ali i njegov radni okvir za izradu dinamičnih veb sajtova pod imenom Django.

## **1.1 Ideja rada i cilj rada**

Ideja rada je pronalaženje odgovarajućeg skupa podataka nad kojim će se vršiti manipulacija, analiza, vizualizacija podataka i na kraju donošenje zaključaka na osovu rezultata. Sređeni skup podataka bi se zatim koristio kao izvor podataka za kreiranje dinamične veb aplikacije na kojoj bi se prikazivali rezultati analize, statističke vrijednosti i vizualne reprezentacije podataka u vidu grafika/dijagrama. Cilj rada je da se sirovi podaci iz skupa podataka prikažu na interfejsu veb aplikacije. Ova aplikacija bi omogućila korisniku da vidi samo one podatke koji su njemu interesantni i značajni za donošenje zaključaka.

## **1.2 Očekivanja od rada**

Očekivanja su da praktični dio ovog rada predstavlja spoj dvije discipline u IT industriji, i to razvoja veb aplikacija i nauke o podacima. A od ukupnog istraživačkog rada (teorijski i praktični dio) se očekuje da donese novinu u ove dvije oblasti, tj. da pokrene dalji razvoj ideja na ovu temu.

## **1.3 Tema u okviru mreže međuzavisnosti**

Šta treba ovdje da se piše?

# **2. Metodologija**

U ovom poglavlju je naveden materijal i metodologija korišćena za izradu praktičnog dijela projekta. Praktični dio projekta je kao što je ranije navedeno veb aplikacija koja prikazuje rezultate analize podataka i vizualne reprezentacije podataka (grafike). Ovo poglavlje je podijeljeno na dva potpoglavlja, i to prvo potpoglavlje u kojem su opisane biblioteke korišćene za projekat, i drugo u kome je opisan skup podataka koji je korišćen.

## **2.1 Softverski paketi**

U ovom poglavlju su detaljno opisani paketi koji su korišćeni za potrebe realizacije praktičnog dijela ovog rada. Paketi koji su korišćeni su:

1. NumPy – je izuzetno brz i jednostavan paket za manipulaciju nad višedimenzionalnim nizovima, vektorima i matricama. „NumPy kombinuje moć programiranja nizova, performanse C-a, čitljivost i svestranost Python-a u dobro testiranoj, dokumentovanoj i zreloj biblioteci za korišćenje“. (Harris, C. R.; Millman, K. J. (2020). Array programming with NumPy. Nature, strana br. 361) Kao što je navedeno NumPy ima brzo izvršavanje poput C programskog jezika koji je po tome poznat. Samim tim nije ni čudno što je većina biblioteka koje slijede napravljeno upravo sa NumPy paketom u osnovi. Ovaj paket nije direktno korišćen u značajnoj mjeri kao ostali paketi, ali jeste indirektno kao njihov sastavni dio.
2. Pandas – je jednostavan i popularan Python softverski paket koji se koristi u analizi i manipulaciji nad podacima. Pandas uvodi dvije vrste novih objekata, i to DataFrame objekte kao dvodimenzionalne, i Series objekte kao jednodimenzionalne strukture. Kao što navodi McKinney: DataFrame objekat se sastoji od većeg broja Series objekata, pa se može reći da su oni u odnosu tabela i kolona. (McKinney, W. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, strana br. 60) Pandas je u praktičnom dijelu korišćen za čišćenje, manipulisanje i analiziranje podataka iz skupa podataka, koji je u vidu CSV fajla. Pandas je korišćen i u dijelu projekta koji se bavio analiziranjem skupa podataka, ali je korišćen i na veb aplikaciji.
3. Matplotlib – je paket koji se koristi za vizualizaciju podataka. Sa ovim paketom je moguće kreirati veliki broj grafika (pita dijagrami, dijagrami sa stubićima itd.). Matplotlib može da radi sa Python listama, NumPy nizovima, ali i iz prethodno pomenutim Pandas objekatima (DataFrame, Series). Ovaj paket je u radu korišćen za vizualni prikaz podataka i u dijelu analize, a takođe i na veb aplikaciji.
4. Seaborn – je takođe paket za vizualizaciju podataka. Razlika između Matplotlib-a i Seaborn-a je kako navodi Michael L. Waskom u tome što: Matplotlib predstavlja paket nižeg nivoa, pa je sa Seaborn paketom kompleksnije statističke grafike mnogo jednostavnije predstaviti nego sa Matplotlib-om. (Waskom, M. L. (2021). Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021. strana br. 1) Seaborn je u projektu korišćen za prikazivanje atraktivnih i kompleksnijih dijagrama kako u analizi, tako i u izradi veb sajta.
5. Django -

# **LITERATURA**

1. Barnett, T.; Jain, S. (2018). Cisco visual networking index (vni) complete forecast update, 2017–2022. Americas/EMEAR Cisco Knowledge Network (CKN) Presentation, 1-30.
2. Harris, C. R.; Millman, K. J. (2020). Array programming with NumPy. Nature.
3. McKinney, W. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference.
4. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering.
5. Waskom, M. L. (2021). Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021.