

**SVEUČILIŠTE U ZAGREBU**  
**FAKULTET ORGANIZACIJE I INFORMATIKE**  
**V A R A Ž D I N**

**Jan Malnar**

**Integracija alata Tableau i programskog  
jezika Python**

**ZAVRŠNI RAD**

**Varaždin, 2023.**

**SVEUČILIŠTE U ZAGREBU**  
**FAKULTET ORGANIZACIJE I INFORMATIKE**  
**V A R A Ž D I N**

**Jan Malnar**

**JMBAG: 0016149145**

**Studij: Primjena informacijske tehnologije u poslovanju**

**Integracija alata Tableau i programskog jezika Python**  
**ZAVRŠNI RAD**

**Mentor:**

Prof. dr. sc. Kornelije Rabuzin

**Varaždin, 2023.**

*Jan Malnar*

### **Izjava o izvornosti**

Izjavljujem da je moj završni rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristio drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

*Autor potvrdio prihvaćanjem odredbi u sustavu FOI-radovi*

---

## **Sažetak**

U ovome završnom radu biti će opisani pojmovi poslovne inteligencije kao i njenih prednosti i nedostataka. Programski jezik Python i alat Tableau biti će opisani i prikazani na praktičnom primjeru usporedbe i vizualizacije podataka o cijenama goriva za 2022. godinu u Hrvatskoj, Mađarskoj i Sloveniji koji će se prethodno očistiti i spremiti za obradu u okruženju Juypiter Notebook gdje će se koristiti programski jezik Python u integraciji sa alatom Tableau pomoću kojeg će biti vizualno prikazani. Cilj ovoga rada je prikazati način, metode i alate za obradu i vizualizaciju podataka koji poduzećima pomažu u poslovnom odlučivanju.

**Ključne riječi:** poslovna inteligencija; obrada podataka; analiza podataka; vizualizacija podataka; Python; Tableau;

# Sadržaj

Sadržaj .....	iii
1. Uvod .....	1
2. Poslovna inteligencija.....	2
2.1. Koncept Poslovne Inteligencije .....	2
2.2. Vrste Alata i Tehnologija .....	2
2.3. Važnost za Strateško Donošenje Odluka.....	3
2.4. Implementacijski Izazovi.....	3
2.5. Budućnost Poslovne Inteligencije .....	3
2.5.1. Alati Jupyter Notebook i Tableau .....	3
2.5.2. Povijest i razvoj .....	4
2.5.3. Struktura i funkcionalnosti .....	4
3. Opis korištenih podataka .....	5
3.1. Sadržaj podataka.....	5
3.2. Informacije o podacima.....	6
4. Analiza i obrada podataka .....	8
4.1. Radno okruženje alata Jupyter Notebook.....	8
4.2. Priprema podataka za obradu .....	10
4.3. Promjena tipa podataka .....	11
4.4. Prosječna, minimalna i maksimalna cijena goriva .....	12
4.5. Minimalna i maksimalna cijena prema gorivu .....	13
4.6. Promjena cijena .....	14
4.7. Korelacijska matrica .....	15
5. Vizualizacija i interpretacija.....	17
5.1. Radno okruženje alata Tableau .....	17
5.2. Vizualizacija prosječne cijene .....	18
5.3. Interpretacija prosječnih cijena.....	20
5.4. Vizualizacija minimalne i maksimalne cijene .....	20
5.5. Interpretacija minimalne i maksimalne cijene.....	21
5.6. Promjena cijena po državi .....	23
5.7. Interpretacija promjene cijene prema vrsti goriva.....	24
5.8. Vremenska serija cijena.....	26
5.9. Interpretacija vremenske serije cijena .....	27
5.10. Korelacijska matrica .....	28

5.11.	Interpretacija korelacijske matrice .....	29
5.12.	Razlika cijene i poreza .....	31
5.13.	Interpretacija razlike cijene goriva i poreza .....	33
6.	Zaključak .....	34
	Popis literature.....	35
	Popis slika .....	36
	Prilozi .....	37

# 1. Uvod

Ovaj završni rad se bazira na pojmu poslovne inteligencije koji će kasnije biti detaljno opisan. Često se koristi i termin BI (eng. Business Intelligence). S obzirom da danas svaka veća firma zahtjeva obradu velikih količina podataka, stoga je potrebno imati stručnjake koji će omogućiti interpretiranje korisnih informacija za poduzeće. Ovaj završni rad ima za cilj pružiti detaljniji pregled pojma poslovne inteligencije na primjeru obrade stvarnih podataka. Metodologija ovoga završnog rada uključuje prikupljanje podataka, obradu i analizu te naposljetku vizualizaciju.

Kao realan primjer biti će prikazana obrada podataka, usporedba i vizualizacija podataka pomoću programskog jezika Python te alata za vizualizaciju Tableau. Podaci su stvarne tjedne cijene goriva za 2022. godinu prema vrsti goriva, datumu te državi u kojoj se nalaze. Podaci koji će se obrađivati vezani su samo uz cijene u Hrvatskoj, Sloveniji i Mađarskoj.

Podaci koji će se obrađivati sadrže informacije o cijenama goriva prema vrsti i datumu. Vrste goriva prema podacima koji su preuzeti sa europskog povjerenstva za energetiku su podijeljeni na dizelsko gorivo (engl. *Automotive gas oil*), Euro-super 95, Lož ulje sumpor do 1% (engl. *Fuel oil – Sulphur less than 1%*), lož plinsko ulje (engl. *Heating gas oil*), ukapljeni naftni plin LPG. Cijene su izražene u eurima u količini od 1000 litara.

Programiranje u Pythonu biti će u Jupyter Notebook okruženju koji olakšava obradu i analizu podataka te će se za vizualizaciju koristiti alat Tableau. Navedeni alati pomažu u otkrivanju trendova, obrazaca i značajki koji mogu pomoći u razumijevanju tržišta te donositeljima odluka u poslovnom svijetu.

## **2. Poslovna inteligencija**

U ovom poglavlju treba opisati koje će metode i tehnike biti korištene pri razradi teme, kako su provedene istraživačke aktivnosti, koji su programski alati ili aplikacije korištene.

Poslovna inteligencija (BI) nije samo alat ili tehnologija; to je strateški resurs koji omogućuje organizacijama da obrade velike količine podataka i iz njih izvuku značajne informacije. U posljednjem desetljeću, BI je postao integralni dio poslovnog ekosustava, koji omogućava strateško donošenje odluka i operativnu izvrsnost (Chen, Chiang & Storey, 2012).

### **2.1. Koncept Poslovne Inteligencije**

Na osnovnoj razini, poslovna inteligencija se odnosi na korištenje podataka za bolje razumijevanje tržišnih trendova, ponašanja potrošača, i internih procesa u organizaciji. Svrha BI je omogućiti tvrtkama da pretvore sirove podatke u korisne informacije koje mogu koristiti za planiranje, optimizaciju i inovacije.

### **2.2. Vrste Alata i Tehnologija**

U kontekstu BI, postoji mnogo alata i tehnologija koje organizacije koriste. Od tradicionalnih relacijskih baza podataka do naprednih algoritama strojnog učenja, raspon je iznimno širok. To uključuje tehnologije za prikupljanje podataka, sustave za izvješćivanje, analitičke platforme i mnoge druge. Jedan od najpoznatijih alata u ovom kontekstu je Tableau, koji omogućuje korisnicima da jednostavno stvore složene vizualizacije i izvještaje. Ako uz to još kombiniramo i neki od programskih jezika pomoću kojih možemo istražiti, proanalizirati i obraditi podatke možemo doći do korisnih informacija za određeno poduzeće ukoliko ih pravilno interpretiramo.



## **2.3. Važnost za Strateško Donošenje Odluka**

Jedna od ključnih prednosti poslovne inteligencije je njezina sposobnost da informira strateško donošenje odluka. Kako Gartner navodi, "BI je ključni element u donošenju informiranih odluka poslovanja" (Gartner, 2019). Kroz analizu podataka, BI omogućuje menadžerima i donositeljima odluka da vide obrasce, identificiraju prilike ili prijetnje, i donose odluke koje su temeljene na stvarnim informacijama, a ne samo na intuiciji ili pretpostavkama.

## **2.4. Implementacijski Izazovi**

Unatoč mnogim prednostima, implementacija BI tehnologija nije bez izazova. Ovi izazovi mogu biti tehnički, kao što su pitanja integracije i interoperabilnosti, ali i organizacijski, uključujući otpor prema promjeni i prilagodba korporativne kulture. Da bi BI bio uspješan, potrebna je sinergija između tehnologije, ljudi i procesa.

## **2.5. Budućnost Poslovne Inteligencije**

U budućnosti, očekuje se da će BI nastaviti evoluirati s rastom novih tehnologija kao što su umjetna inteligencija i Internet stvari (IoT), Cloud BI. Ovi napredci će dodatno proširiti mogućnosti za analizu podataka i stvaranje novih poslovnih modela, pružajući tvrtkama dodatne prilike za rast i inovacije.

### **2.5.1. Alati Jupyter Notebook i Tableau**

Jupyter Notebook predstavlja revolucionarno okruženje za izvođenje interaktivne analize podataka i razvoj koda. Njegova sposobnost da kombinira kod, tekst, matematičke formule i vizualizacije u jednom dokumentu omogućuje sveobuhvatni pristup analizi podataka. U kontekstu ovog istraživanja, Jupyter Notebook će se koristiti u kombinaciji s programskim jezikom Python i bibliotekom Pandas za dubinsku analizu cijena goriva u Hrvatskoj, Sloveniji i Mađarskoj.

Nakon obrade podaci će na vizualan i privlačan način biti jasno prikazani što je u poslovnoj inteligenciji vrlo važno zbog dubljeg uvida u podatke i lakših donošenja odluka stoga

je u ovome završnom radu korišten alat za vizualizaciju podataka Tableau. Tableau je alat koji se koristi za pretvaranje podataka u razumljive i još važnije ima mogućnost vizualne interakcije. Masovno je korišten u polju poslovne inteligencije. Jednostavan je za korištenje, podržava mogućnost korištenja velikih količina podataka te ima mogućnost integracije sa raznim izvorima podataka kao što su SQL baze podataka, cloud servisi, Excel i ostali...

### **2.5.2. Povijest i razvoj**

Jupyter Notebook je projekt koji je započeo kao IPython (Interactive Python) 2001. godine, a od tada je proširio svoje mogućnosti na mnoge druge programske jezike. Ideja iza njega bila je stvoriti interaktivno računalno okruženje u kojem korisnici mogu brzo i jednostavno izvoditi niz računalnih zadataka. Projekt je doživio brzi razvoj i danas je postao standardni alat u domeni podatkovne znanosti. Stoga je ovo okruženje korišteno u ovome završnom radu.

Tableau Software osnovan je 2003. godine u Sjedinjenim Američkim Državama. Poduzeće je osnovano s ciljem omogućavanja korisnicima da bolje vizualiziraju i razumiju svoje podatke. Tableau je brzo stekao reputaciju kao jedan od vodećih alata za vizualnu analizu podataka. Tableau danas ostaje jedan od najpoznatijih alata za BI i vizualnu analizu podataka, koristi se u mnogim industrijama i nastavlja biti na čelu inovacija u području poslovne inteligencije.

### **2.5.3. Struktura i funkcionalnosti**

Jupyter Notebook se sastoji od niza ćelija koje mogu sadržavati tekst, kod ili druge elemente. Tekstualne ćelije koriste Markdown za formatiranje, dok se kod ćelije izvode u izoliranom okruženju zvanom "kernel". Jedna od glavnih prednosti Jupyter Notebooka je mogućnost izravnog izvođenja koda i pregleda rezultata unutar samog dokumenta, što znatno olakšava iterativni razvoj i analizu. U nastavku ovog završnog rada na stvarnom primjeru biti će opisane način obrade podataka te naposljetku njihova vizualizacija.

U ovom završnome radu koristiti će se Tableau Desktop koji omogućuje korisnicima da kreiraju interaktivne i djeljive sadržaje. Korištenje je vrlo intuitivno na način da se mogu povući i ispustiti podaci koji su objekt promatranja na osi te je moguća izrada raznih vrsta grafova i tablica. Posjeduje mogućnosti naprednih opcija za filtriranje, sortiranje i grupiranje podataka te ih je naposljetku moguće izvesti u različitim formatima poput PDF-a, PNG i tomu slično. Osim toga jedna od njegovih prednosti je ta što se može koristiti u integraciji sa ostalim alatima ili platformama poput programskog jezika R i Pythona.

### 3. Opis korištenih podataka

Korišteni podaci preuzeti su sa stranice Europske komisije koja sadrži podatke o potrošačkim cijenama naftnih derivata za svaki tjedan počevši od 1995 pa sve do danas i u narednom periodu. Nacionalni podaci obično se komisiji dostavljaju srijedom. U ovome završnom radu koristiti će se podaci o cijenama naftnih derivata u Hrvatskoj, Sloveniji i Mađarskoj u 2022. godini.

#### 3.1. Sadržaj podataka

	Prices in force on	Country Name	Country EU Code	Product Name	Currency Code	Prices Unit	Euro exchange rate	Weekly price with taxes	Weekly price without taxes
0	2022-01-03	Croatia	HR	Automotive gas oil	HRK	1000L	0.132996	1,463.49	763.82
1	2022-01-03	Croatia	HR	Euro-super 95	HRK	1000L	0.132996	1,484.9	674.56
2	2022-01-03	Croatia	HR	Fuel oil - Sulphur less than 1%	HRK	t	0.132996	712.73	691.45
3	2022-01-03	Croatia	HR	Heating gas oil	HRK	1000L	0.132996	756.75	559.78
4	2022-01-03	Croatia	HR	LPG - motor fuel	HRK	1000L	0.132996	786.27	621.65

Slika 1: Prikaz tablice s podatcima (izvor: vlastita izrada)

Na slici 1 možemo vidjeti tablicu sa podacima cijena naftnih derivata. Podaci se sastoje od devet stupaca. Najprije imamo stupac „Prices in force on“ koji predstavlja datum kojim se označava da na taj datum cijena stupa na snagu. Stupac „Country Name“ predstavlja ime države u kojoj cijena stupa na snagu, u ovom slučaju biti će prikazane cijene samo za prethodno navedene tri države zbog velike količine podataka. Stupac „Country EU Code“ predstavlja kraticu za državu koja se nalazi u prethodnom stupcu. Stupac „Product Name“ predstavlja vrstu ili kategoriju naftnog derivata u određenoj državi. Stupac „Currency Code“ označava nacionalnu valutu države u kojoj cijene stupaju na snagu. Stupac „Price unit“ označava mjernu jedinicu za količinu naftnih derivata te je iskazana u litrama ili na ponekim mjestima u tonama, no kasnije u nastavku završnog rada sve će biti iskazano u litrama transformacijom podataka. Stupac „Euro exchange rate“ predstavlja tečaj eura koji se odnosi na tečaj američkog dolara jer se obično cijena naftnih derivata u svijetu prikazuje u američkim dolarima. Stupac „Weekly price with taxes“ prikazuje tjednu cijenu naftnog derivata sa porezom dok stupac „Weekly price without taxes“ prikazuje cijenu naftnih derivata bez poreza. Ovakav

prikaz podataka predstavlja sirove neobrađene podatke koje je potrebno da se dodatno obrade te da podaci budu spremni i prikladni za daljnju analizu.

## 3.2. Informacije o podacima

Prije provođenja analize i obrade podataka, bitno je temeljito razumjeti različite karakteristike tih podataka. To uključuje informacije kao što su tip podatka, indeksi te prostor zauzimanja memorije, jer su to sve ključni elementi za uspješno manipuliranje i interpretaciju podataka. Tablica sa prethodne slike koja prikazuje podatke je jedan „DataFrame“ koji predstavlja osnovnu struktura podataka u Pythonovoj biblioteci Pandas, koja se široko koristi za analizu podataka. Pojednostavljeno, može se smatrati tablicom koja se sastoji od redaka i stupaca, slično kao Excelova radna tablica ili relacijska baza podataka. Svaki stupac u DataFrame-u može biti različitog tipa podataka (npr. cijeli brojevi, decimalni brojevi, tekst, datumi itd.). U nastavku prilikom korištenja Python funkcije info u Jupyter Notebook-u dobijemo detaljnije informacije o podacima:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 714 entries, 0 to 713
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Prices in force on                    714 non-null    datetime64[ns]
1   Country Name                          714 non-null    object
2   Country EU Code                       714 non-null    object
3   Product Name                          714 non-null    object
4   Currency Code                         714 non-null    object
5   Prices Unit                           714 non-null    object
6   Euro exchange rate                    714 non-null    float64
7   Weekly price with taxes                714 non-null    object
8   Weekly price without taxes             714 non-null    object
dtypes: datetime64[ns](1), float64(1), object(7)
memory usage: 50.3+ KB
```

Ako promotrimo prethodno dobivene informacije o podacima možemo uočiti sljedeće stavke koje su korisne za daljnju obradu i analizu podataka:

- **<class 'pandas.core.frame.DataFrame'>**: Označava da je podatkovni skup tipa DataFrame u Pandas-u.
- **RangeIndex: 714 entries, 0 to 713**: Ukupno ima 714 unosa (redaka) u DataFrame-u, indeksirano od 0 do 713.
- **Data columns (total 9 columns)**: Ukupno ima 9 stupaca u DataFrame-u.

- **Non-Null Count:** Za svaki stupac, ovo označava broj unosa koji nisu null (prazni ili nedostajući). Svi stupci imaju 714 unosa koji nisu null, što znači da nema nedostajućih podataka.
- **Dtype:** Ovo označava tip podatka za svaki stupac. Na primjer, "Prices in force on" je tipa **datetime64[ns]**, što znači da je u formatu datuma i vremena. "Euro exchange rate" je **float64**, što znači da su to decimalni brojevi. Ostali su **object**, što obično znači da su tekstualni ili mješoviti tipovi podataka.
- **memory usage: 50.3+ KB:** Ukupna količina memorije koju DataFrame koristi je nešto više od 50.3 KB.

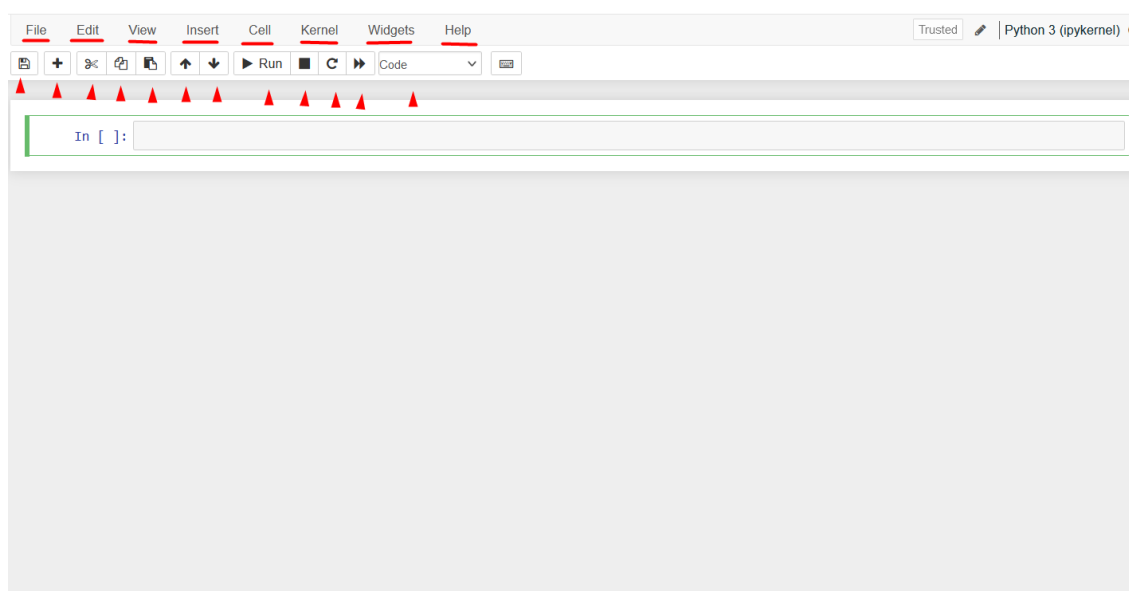
S obzirom da nam treba brojčani tip podataka za cijene prije nego započnemo s daljnom analizom biti će potrebno pretvoriti vrijednosti cijena naftnih derivata „Weekly price with taxes“ i „Weekly price without taxes“ iz tipa podataka object u float kako bi bili sigurni da su vrijednosti decimalne vrijednosti.

## 4. Analiza i obrada podataka

Analiza i obrada prethodno opisanih podataka biti će ostvarena pomoću programskog jezika Python u okruženju Jupyter Notebook. U nastavku biti će prikazano radno okruženje Jupyter Notebook-a koje je pokrenuto pomoću Anaconda Prompt naredbenog retka koji omogućava pokretanje Jupyter Notebook sesije pomoću komande `jupyter notebook`.

### 4.1. Radno okruženje alata Jupyter Notebook

Na sljedećoj slici prikazano je radno okruženje alata Jupyter Notebooka u kojem ćemo provoditi obradu i analizu podataka pomoću programskog jezika Python.



**Slika 2:** Prikaz radnog okruženja alata Jupyter Notebook (izvor: vlastita izrada)

Na slici 2 možemo vidjeti izbornike radnog okruženja podcrtane crvenom bojom, pa tako imamo sljedeće mogućnosti na prethodnoj slici:

- **File** – izbornik koji sadrži osnovne opcije za upravljanje notebookom, kao što su stvaranje novih notebooka, otvaranje postojećih, spremanje i ispis.
- **Edit** – Izbornik "Edit" omogućuje manipulaciju ćelijama unutar notebooka. Pa je tako moguće kopirati, izrezati i zalijepiti ćelije, kao i pronaći i zamijeniti tekst.
- **View** – U izborniku "View" moguće je prilagoditi kako notebook izgleda. Ovdje se može uključiti ili isključiti zaglavlje i alatnu traku.

- **Insert** – Izbornik "Insert" koristi se za dodavanje novih ćelija iznad ili ispod trenutačno odabrane ćelije.
- **Cell** – Izbornik "Cell" pruža opcije za izvršavanje ćelija, mijenjanje tipa ćelija i druge akcije koje se odnose na ćelije.
- **Kernel** – izbornik koji omogućuje upravljanje jezgrom (kernel) koji pokreće notebook. Moguće je pokrenuti jezgru, prekinuti izvršavanje i ponovno se povezati ako je potrebno.
- **Widgets** – Izbornik "Widgets" služi za upravljanje interaktivnim widgetima unutar notebooka, ako su instalirani.
- **Help** – "Help" izbornik pruža pristup korisničkom sučelju, tipkovničkim prečacima i dokumentaciji kako biste lakše koristili Jupyter Notebook.

Ispod izbornika imamo na prethodnoj slici imamo ikonice koje su označene crvenim strelicama pa su one opisane po redu s lijeva na desno:

- **Save** – Prva ikona diska koja koristi za spremanje trenutačnog stanja notebooka.
- **+** - Ikonica „*Insert cell below*“ služi za dodavanje nove ćelije ispod trenutno odabrane ćelije.
- **Škare, copy, paste** – tri ikonice koje služe za izrezivanje, kopiranje i lijepljenje odabrane ćelije.
- **Strelice prema gore i dolje** – Ikonice služe za pomicanje odabranih ćelija na prema gore ili dolje.
- **Play** – Ikonica koja služi za izvršavanje trenutačno odabrane ćelije i pomicanje fokusa na sljedeću ćeliju.
- **Stop** – Ikonica koja služi za prekid izvršavanja ćelije.
- **Restart kernel** – Ikonica kružne strelice koristi se za ponovno pokretanje jezgre (kernel-a).
- **Cell type** – Dropdown izbornik koji se koristi za promjenu tipa ćelije kao na primjer (kod, markdown, naslov).
- **Command Palette** – Ikonica sa gumbom koja tipkovnice koja otvara izbornik sa definiranim Jupyter naredbama koje je moguće pretražiti i izvršiti prema vlastitim potrebama.

## 4.2. Priprema podataka za obradu

Na sljedećoj slici uočiti ćemo da ćemo koristiti Pandas Python biblioteku koja služi za analizu i manipulaciju podacima. Potom će biti učitani podaci iz Excel tablice te će biti ispisanih prvih par redova podataka za bolji pregled.

```
In [36]: import pandas as pd
In [37]: df = pd.read_excel('E:/cijene/Cijene.xlsx')
In [38]: df.head()
```

Out[38]:

	Prices in force on	Country Name	Country EU Code	Product Name	Currency Code	Prices Unit	Euro exchange rate	Weekly price with taxes	Weekly price without taxes
0	2022-01-03	Croatia	HR	Automotive gas oil	HRK	1000L	0.132996	1,463.49	763.82
1	2022-01-03	Croatia	HR	Euro-super 95	HRK	1000L	0.132996	1,484.9	674.56
2	2022-01-03	Croatia	HR	Fuel oil - Sulphur less than 1%	HRK	t	0.132996	712.73	691.45
3	2022-01-03	Croatia	HR	Heating gas oil	HRK	1000L	0.132996	756.75	559.78
4	2022-01-03	Croatia	HR	LPG - motor fuel	HRK	1000L	0.132996	786.27	621.65

Slika 3: Prikaz učitavanja podataka i biblioteke Pandas (izvor: vlastita izrada)

Na slici 3 je vidljivo da najprije je učitana Pandas biblioteka u prvoj ćeliji pomoću funkcije **import** te joj je dodijeljen alias **pd**. Nakon učitane biblioteke definirana je varijabla **df** koja označava dataframe koji je prethodno opisan te su u njega učitani podaci iz Excel tablice pomoću metode **pd.read\_excel()** koja čita podatke iz tablice i pohranjuje ih u dataframe. Nakon učitanih podataka ispisani su prvih pet redaka dataframe-a pomoću metode **head()** pomoću koje možemo dobiti bolji uvid u podatke te ih obrađivati i po potrebi transformirati.

U ovom slučaju vidljivo je iz slike da imamo devet stupaca sa podacima koji su prethodno opisani te njihove tipove. Treba primijetiti da na nekim mjestima u stupcu „Prices Unit“ koji označava mjernu jedinicu goriva je napisana u mjernoj jedinici t što označava tonu stoga je bitno da podaci budu pregledniji te će biti potrebno u nastavku promijeniti sve mjerne jedinice iz t u 1000L te je potrebno kao što je prethodno navedeno promijeniti tipove podataka „Weekly price with taxes“ i „Weekly prices without taxes“ iz tipa podataka **object** u tip podatka **float** kako bi bili sigurni da se koriste brožčane vrijednosti za cijene goriva.



## 4.3. Promjena tipa podataka

Na sljedećoj slici vidljiv je programski kod koji će prebaciti stupce s cijenama u tip podatka **float** te ukloniti zareze u odvajanju tisućica na sljedeći način:

```
In [40]: #Najprije je potrebno prebaciti stupac sa cijenama u float te ukloniti zareze u odvajanju tisućica
df['Weekly price with taxes'] = df['Weekly price with taxes'].str.replace(',', '').astype(float)
df['Weekly price without taxes'] = df['Weekly price without taxes'].str.replace(',', '').astype(float)
```

Slika 4: Prikaz pretvorbe tipa podataka (izvor: vlastita izrada)

**df** ["Weekly price with taxes"] se odnosi na stupac pod imenom "Weekly price with taxes" u DataFrame-u **df**. Nakon toga koristi se **str** funkcija koja omogućava metode za manipulaciju stringa (znakovni tip podatka) kao na primjer u ovom slučaju koristi metodu **replace** koja zamjenjuje sve zareze ',' u prazan string ' ' te time numeričke vrijednosti predstavljene kao string pretvaramo u pravilni numerički format na primjer broj „1,000“ će postati broj „1000“. Na posljepku se mijenja tip podatka odabranog stupca u DataFrame-u pomoću funkcije **astype()** u tip **float**. Isti postupak je ponovljen i za stupac u dataframe-u **df** ["Weekly price without taxes"] koji se odnosi na cijene bez poreza.

```
In [41]: #Zamjena mjerne jedinice tako da svuda piše 1000L
df['Prices Unit'] = df['Prices Unit'].replace({'t': '1000L'})
df.to_excel("E:/Podaci za tableau/Cijenetransofrmirano.xlsx", index=False)
```

Slika 5: Prikaz zamjene mjernih jedinica. (izvor: vlastita izrada)

Na prethodnoj slici 5 u stupcu u dataframe-u **df**["Prices Unit"] također je primjenjena metoda **replace()** kojom su zamijenjeni stringovi mjerne jedinice 't' u '1000L' kako bi mjerne jedinice bile kontinuirano jednake i preglednije. Nakon toga korištena je metoda **to\_excel()** koja sprema dataframe u Excel datoteku koja će kasnije biti korištena za vizualizaciju cijena u alatu Tableau. **Index=False** označava da indeks stupca koji smo vidjeli u prethodnom ispisu tablice dataframe-a neće biti spremljen u Excel datoteku. Također važno je i napomenuti da je potrebno navesti putanju i naziv Excel datoteke gdje će se spremiti podaci. Na sljedećoj slici možemo vidjeti prethodno navedene promjene u DataFrame-u:

```
In [42]: df.head()
```

Out[42]:

	Prices in force on	Country Name	Country EU Code	Product Name	Currency Code	Prices Unit	Euro exchange rate	Weekly price with taxes	Weekly price without taxes
0	2022-01-03	Croatia	HR	Automotive gas oil	HRK	1000L	0.132996	1463.49	763.82
1	2022-01-03	Croatia	HR	Euro-super 95	HRK	1000L	0.132996	1484.90	674.56
2	2022-01-03	Croatia	HR	Fuel oil - Sulphur less than 1%	HRK	1000L	0.132996	712.73	691.45
3	2022-01-03	Croatia	HR	Heating gas oil	HRK	1000L	0.132996	756.75	559.78
4	2022-01-03	Croatia	HR	LPG - motor fuel	HRK	1000L	0.132996	786.27	621.65

Slika 6 Prikaz DataFrame nakon promjene. (izvor: vlastita izrada)

## 4.4. Prosječna, minimalna i maksimalna cijena goriva

U nastavku će biti opisan način izračuna prosječne, minimalne i maksimalne cijene goriva za svaku državu i za svaku vrstu goriva. Na sljedećoj slici vidljiv je sljedeći programski kod:

```
In [43]: # Grupiranje dataframe-a prema državi, vrsti goriva, te izračun prosječne, minimalne i maksimalne cijene goriva na određeni datum
# Funkcija za pridruživanje datuma max i min cijenama
def get_max_min_dates(group):
    max_price_idx = group['Weekly price with taxes'].idxmax()
    min_price_idx = group['Weekly price with taxes'].idxmin()
    group['Max na datum'] = group.loc[max_price_idx, 'Prices in force on']
    group['Min na datum'] = group.loc[min_price_idx, 'Prices in force on']
    return group

# Grupiranje i primjena funkcije
result = df.groupby(['Country Name', 'Product Name'], group_keys=False).apply(get_max_min_dates)

# Izračun prosječne, max i min cijene
result = result.groupby(['Country Name', 'Product Name']).agg({
    'Weekly price with taxes': ['mean', 'max', 'min'],
    'Max na datum': 'first', # Prva vrijednost nakon grupiranja je ona koja je pridružena max cijeni
    'Min na datum': 'first' # Prva vrijednost nakon grupiranja je ona koja je pridružena min cijeni
}).reset_index()

result.columns = ['Country Name', 'Product Name', 'Prosječna cijena', 'Max Cijena', 'Min Cijena', 'Max na datum', 'Min na datum']
result.to_excel("E:/Podaci za tableau/Prosje_max_min_sve_drz.xlsx", index=False)
result
```

Slika 7: Prikaz programskog koda za izračun prosječne, minimalne i maksimalne cijene. (izvor: vlastita izrada)

Najprije je definirana funkcija koja se definira pomoću ključne riječi **def** pod nazivom **get\_max\_min** te joj je dodijeljen grupirani DataFrame kao argument **group** koja predstavlja pojedinačnu grupu podataka koja se izdvaja iz DataFrame-a. U funkciji je definirana varijabla u koju sprema vrijednosti iz stupca „Weekly price with taxes“ pomoću Pandas metode **idxmax()** i **idxmin()** koji vraćaju indeks odabranog retka gdje je maksimalna i minimalna vrijednost. Zatim je korištena metoda **loc[ ]** koja pristupa grupi vrijednosti u ovom slučaju dohvaća se indeks reda iz prethodno definiranih varijabli u funkciji te datum na koji dan je cijena maksimalna „Prices in force on“ te se taj datum dodaje u novi stupac „Max na datum“ i „Min na datum“.

Nakon što je definirana funkcija u varijablu **result** spremaju se rezultati primijenjene funkcije **get\_max\_min\_dates** na grupirane stupce u DataFrame-u „Country Name“ i „Product Name“ gdje se za svaku državu i vrstu goriva primjenjuje prethodno definirana funkcija koja dohvaća datum kada je nastala minimalna i maksimalna cijena goriva za svaku državu i za svaku vrstu goriva te je spriječeno dodavanje grupnih ključeva u indeks pomoću metode **group\_keys=False**. Nakon grupiranja podatak korištena je funkcija agregacije **agg()** koja omogućava primjenu niza operacija nad grupiranim podacima pa tako je izračunata prosječna, maksimalna i minimalna cijena pomoću metoda **mean**, **max** i **min**. Te se naposljetku izvlači prvi datum metodom **first** koji odgovara minimalnoj i maksimalnoj cijeni, time je omogućeno da se sažmu podaci i budu spremni ukoliko budu bili potrebni za daljnju analizu i vizualizaciju.

Te je na kraju promijenjen naziv stupaca za bolju preglednost pa su tako na sljedećoj slici vidljivi rezultati prethodno opisanog koda.

	Country Name	Product Name	Prosječna cijena	Max Cijena	Min Cijena	Max na datum	Min na datum
0	Croatia	Automotive gas oil	1762.536078	1947.25	1463.49	2022-07-04	2022-01-03
1	Croatia	Euro-super 95	1604.532549	1873.26	1331.17	2022-06-06	2022-12-26
2	Croatia	Fuel oil - Sulphur less than 1%	922.257647	1110.35	711.89	2022-03-21	2022-01-10
3	Croatia	Heating gas oil	1188.309200	1471.53	756.75	2022-07-11	2022-01-03
4	Croatia	LPG - motor fuel	944.967255	1050.85	786.27	2022-06-13	2022-01-03
5	Hungary	Automotive gas oil	1478.166078	1778.38	1220.32	2022-12-26	2022-03-07
6	Hungary	Euro-super 95	1289.545882	1585.74	1202.99	2022-12-19	2022-10-10
7	Hungary	Fuel oil - Sulphur less than 1%	815.240980	960.96	559.97	2022-06-06	2022-01-24
8	Hungary	Heating gas oil	1478.166078	1778.38	1220.32	2022-12-26	2022-03-07
9	Hungary	LPG - motor fuel	954.940000	1022.27	839.89	2022-08-08	2022-03-07
10	Slovenia	Automotive gas oil	1635.272353	1865.51	1367.28	2022-07-04	2022-01-03
11	Slovenia	Euro-super 95	1493.366471	1772.32	1318.03	2022-07-04	2022-01-03
12	Slovenia	Heating gas oil	1326.694706	1672.12	974.76	2022-06-27	2022-01-03
13	Slovenia	LPG - motor fuel	975.042549	1054.42	833.37	2022-04-25	2022-02-07

**Slika 8:** Prikaz DataFrame-a nakon izračuna. (izvor: vlastita izrada)

Na slici su vidljivi rezultati te su prikazane prosječne, minimalne i maksimalne cijene za svaku vrstu goriva u svakoj državi. Svaka vrijednost je nastala na određeni datum stoga je prikazan datum nastajanja maksimalne i minimalne cijene.

## 4.5. Minimalna i maksimalna cijena prema gorivu

Na sljedećoj slici prikazan je programski kod koji računa minimalnu i maksimalnu cijenu goriva prema vrsti goriva, pa će se tako moći lakše uočiti koje gorivo je najskuplje ili najjeftinije te u kojoj državi:

```
In [44]: # Izračun minimalne i maksimalne cijene prema vrsti goriva
max_prices = df.groupby('Product Name')['Weekly price with taxes'].idxmax()
min_prices = df.groupby('Product Name')['Weekly price with taxes'].idxmin()

# Izračun datuma
max_price_df = df.loc[max_prices, ['Product Name', 'Country Name', 'Weekly price with taxes', 'Prices in force on']]
min_price_df = df.loc[min_prices, ['Product Name', 'Country Name', 'Weekly price with taxes', 'Prices in force on']]

max_price_df.columns = ['Product Name', 'Država sa Max Cijenom', 'Max Cijena', 'Datum Max Cijene']
min_price_df.columns = ['Product Name', 'Država sa Min Cijenom', 'Min Cijena', 'Datum Min Cijene']

#Resetiranje indexa
max_price_df.reset_index(drop=True, inplace=True)
min_price_df.reset_index(drop=True, inplace=True)

result = pd.merge(max_price_df, min_price_df, on='Product Name')
result.to_excel("E:/Podaci za tableau/Prosje_max_min_za_goriva.xlsx", index=False)
result
```

**Slika 9:** Prikaz koda za izračun minimalne i maksimalne cijene prema vrsti goriva. (izvor: vlastita izrada)

Na prethodnoj slici vidljivo je da se najprije spremaju vrijednosti u varijable **max\_prices** i **min\_prices**. Te vrijednosti su grupirane prema argumentu „*Product Name*“ te se prema grupiranom stupcu odabiru stupci „*Weekly price with taxes*“ tjedne minimalne i maksimalne cijene slično kao i u prethodnom primjeru. Nakon toga se lociraju datumi prema indeksima koji su pohranjeni u prethodne dvije varijable. Za bolje razumijevanje podataka mijenjaju se imena stupaca pomoću atributa **columns**. Resetiraju se indeksi DataFrame-a gdje **drop=True** odbacuje stare indekse dok **inplace=True** promjene izvršava na originalnom DataFrame-u. Na posljetku se koristi Pandas funkcija **merge()** koja na temelju zajedničkog stupca „*Product Name*“ spaja **max\_price\_df** i **min\_price\_df** varijable te ih se naposljetku ispisuje za bolju preglednost. Na sljedećoj slici su vidljivi rezultati izvršenog koda:

Out[44]:

	Product Name	Država sa Max Cijenom	Max Cijena	Datum Max Cijene	Država sa Min Cijenom	Min Cijena	Datum Min Cijene
0	Automotive gas oil	Croatia	1947.25	2022-07-04	Hungary	1220.32	2022-03-07
1	Euro-super 95	Croatia	1873.26	2022-06-06	Hungary	1202.99	2022-10-10
2	Fuel oil - Sulphur less than 1%	Croatia	1110.35	2022-03-21	Hungary	559.97	2022-01-24
3	Heating gas oil	Hungary	1778.38	2022-12-26	Croatia	756.75	2022-01-03
4	LPG - motor fuel	Slovenia	1054.42	2022-04-25	Croatia	786.27	2022-01-03

Slika 10: Prikaz DataFrame-a sa izračunima maksimalne i minimalne cijene (izvor: vlastita izrada)

Iz tablice se može zamijetiti da su goriva dizel, Euro-super 95 i lož ulje imali najveću cijenu u Hrvatskoj dok su pak najmanju cijenu imali u Mađarskoj. Lož plinsko ulje je imalo najveću cijenu u Mađarskoj dok je pak najmanja cijena bila u Hrvatskoj. U Sloveniji je zapažena maksimalna cijena Ukapljenog naftnog plina dok je najmanja cijena zapažena u Hrvatskoj. Ovakvi podaci iz tablice nisu baš lako uočljivi stoga ih je potrebno vizualizirati što će biti napravljeno u nastavku završnog rada.

## 4.6. Promjena cijena

In [49]:

```
# Sortiranje podataka prema državi, vrsti goriva i datumu
df_promjena = df.sort_values(['Country Name', 'Product Name', 'Prices in force on'])

# Izračunavanje promjena cijena kroz godinu za svaku državu i vrstu goriva
df_promjena['Promjena cijene'] = df_promjena.groupby(['Country Name', 'Product Name'])['Weekly price with taxes'].diff()
```

Slika 11: Prikaz kod za izračun promjene cijena kroz godinu. (izvor: vlastita izrada)

Na slici je vidljiv kod koji izračunava promjenu cijena kroz godinu za svaku državu i za svaku vrstu goriva. Prvi korak u ovom kodu je sortiranje podataka prema državi „*Country Name*“, vrsti goriva „*Product Name*“ i datumu „*Prices in force on*“. Ovo je postignuto korištenjem metode **sort\_values()** na originalnom DataFrame-u **df**. Podaci su uređeni tako da su prvo grupirani prema državi, unutar tih grupa prema vrsti goriva, i naposljetku, unutar tih podgrupa, sortiraju se prema datumu. Rezultat ovog koraka sprema se u novi DataFrame

**df\_promjena.** Nakon sortiranja, kod izračunava tjedne promjene cijena goriva za svaku državu i vrstu goriva. Ovo je postignuto korištenjem metode **groupby()** za grupiranje podataka prema državi i vrsti goriva, a zatim korištenjem metode **diff()** na stupcu „Weekly price with taxes“. Metoda **diff()** izračunava razliku između trenutne i prethodne vrijednosti u svakoj grupi. Ova razlika predstavlja promjenu cijene u odnosu na prethodni tjedan unutar iste države i za istu vrstu goriva. Pa stoga dobijemo sljedeće rezultate kao na sljedećoj slici:

Out[50]:

	Prices in force on	Country Name	Country EU Code	Product Name	Currency Code	Prices Unit	Euro exchange rate	Weekly price with taxes	Weekly price without taxes	Promjena cijene
0	2022-01-03	Croatia	HR	Automotive gas oil	HRK	1000L	0.132996	1463.49	763.82	NaN
5	2022-01-10	Croatia	HR	Automotive gas oil	HRK	1000L	0.132841	1465.50	765.91	2.01
10	2022-01-17	Croatia	HR	Automotive gas oil	HRK	1000L	0.132846	1492.53	787.51	27.03
15	2022-01-24	Croatia	HR	Automotive gas oil	HRK	1000L	0.132820	1519.99	809.56	27.46
20	2022-01-31	Croatia	HR	Automotive gas oil	HRK	1000L	0.132814	1520.19	809.74	0.20

**Slika 12:** Prikaz DataFrame-a sa promjenama cijena kroz godinu. (izvor: vlastita izrada)

Iz tablice može se uočiti stupac koji prikazuje promjenu cijene za svaki tjedan. Treba primijetiti da prva vrijednost ne postoji to je zato što nema prethodne vrijednosti pomoću koje bi se izračunala razlika jer ta vrijednost spada u 2021. godinu koja nije objekt promatranja. Ovako pripremljeni podaci omogućuju analizu kako se cijene goriva mijenjaju tijekom vremena, i to na razini svake države i vrste goriva.

## 4.7. Korelacijska matrica

Korelacijska matrica je kvadratna matrica koja prikazuje korelacijske koeficijente između više varijabli. Svaki element matrice predstavlja korelacijski koeficijent između parova varijabli. Korelacijski koeficijent je vrijednost između -1 i 1 koja odražava jačinu i smjer veze između dviju varijabli. Vrijednost od -1 ukazuje na savršenu negativnu korelaciju, vrijednost od 1 ukazuje na savršenu pozitivnu korelaciju, a vrijednost od 0 ukazuje na nepostojanje korelacije. Korelacijska matrica koristi se u statističkoj analizi za ocjenu odnosa između varijabli. U ovom slučaju gleda se korelacije među cijenama vrsta goriva te se gleda odnos između cijena. Na sljedećoj slici prikazan je kod za korelacijsku matricu cijena goriva:

```
In [59]: grupiranje_po_gorivu = df.groupby(['Product Name', 'Prices in force on'])['Weekly price with taxes'].mean().reset_index()
# Pivotiranje podataka tako da svaka vrsta goriva postane zaseban stupac
pivot_df = grupiranje_po_gorivu.pivot(index='Prices in force on', columns='Product Name', values='Weekly price with taxes')
# Izračun korelacijske matrice
correlation_matrix = pivot_df.corr()
correlation_matrix.to_excel("E:/Podaci za tableau/korelacijska_matrica.xlsx")
correlation_matrix
```

**Slika 13:** Prikaz koda za izračun korelacijske matrice. (izvor: vlastita izrada)

Najprije se grupira DataFrame prema vrsti goriva i datumu te se zatim izračunava prosječna tjedna cijena za goriva, te se resetira indeks DataFrame. Nakon grupiranja potrebno je preoblikovati podatke tako da svaka vrsta goriva ima zaseban stupac dok indeksi postaju datumi te je tako lakše izravno usporediti različite vrste goriva prema datumima te se sprema u varijablu **pivot\_df**. Da bi to uradili potrebno je koristiti metodu **pivot()** te pomoću te metode preoblikujemo podatke. Na poslijetku se primjenjuje metoda **corr()** nad varijablom **pivot\_df** koja izračunava Pearsonov koeficijent korelacije između parova stupaca. Pearsonov koeficijent korelacije mjeri linearnu korelaciju između dva skupa podataka u ovom slučaju rezultat će biti korelacijska matrica koja pokazuje međusobne odnose između različitih goriva. Na sljedećoj slici vidljiv je rezultat korelacijske matrice koji će biti kasnije opisan pomoću vizualizacije u alatu Tableau.

Out[59]:

Product Name	Automotive gas oil	Euro-super 95	Fuel oil - Sulphur less than 1%	Heating gas oil	LPG - motor fuel
Product Name					
Automotive gas oil	1.000000	0.351173	0.786315	0.965548	0.821554
Euro-super 95	0.351173	1.000000	0.483631	0.441623	0.577598
Fuel oil - Sulphur less than 1%	0.786315	0.483631	1.000000	0.803224	0.844689
Heating gas oil	0.965548	0.441623	0.803224	1.000000	0.873595
LPG - motor fuel	0.821554	0.577598	0.844689	0.873595	1.000000

**Slika 14:** Prikaz DataFrame-a korelacijske matrice. (izvor: vlastita izrada)

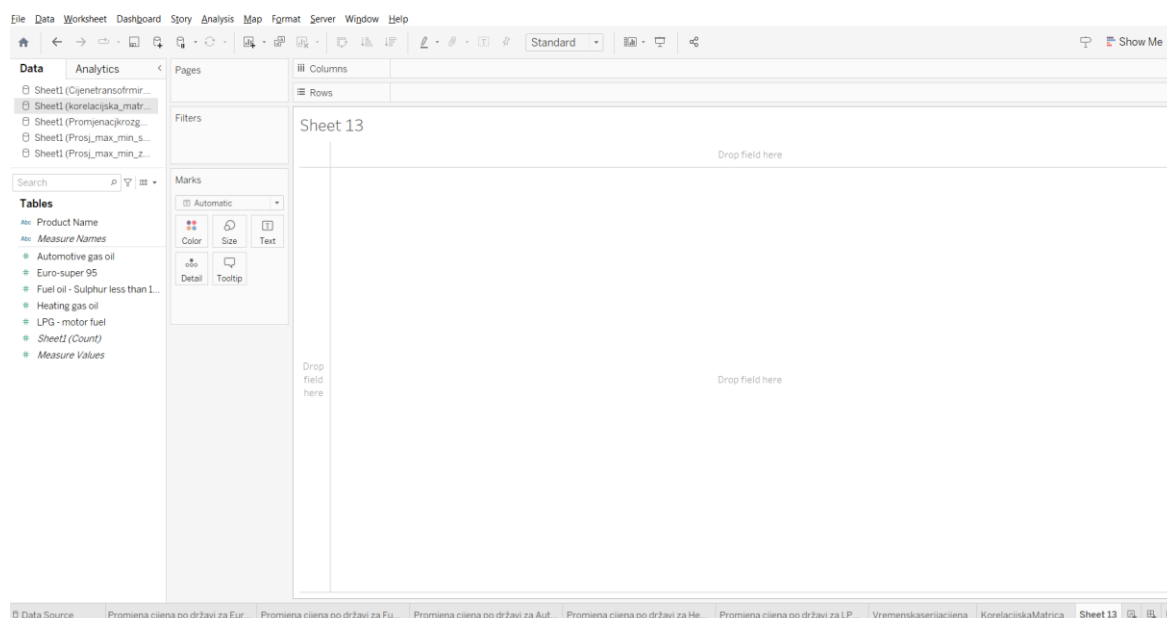
Na slici su vidljivi koeficijenti u rasponu od 0 do 1 koji prikazuju korelacije između cijena vrsta goriva.

## 5. Vizualizacija i interpretacija

Za vizualizaciju podataka koristi se vodeći alat za vizualizaciju Tableau. U nastavku će biti opisano radno okruženje te kako vizualizirati podatke koji su prethodno obrađeni u Pythonu.

### 5.1. Radno okruženje alata Tableau

Na sljedećoj slici prikazano je radno okruženje Tableau-a. Organiziran je tako kako bi omogućilo intuitivno i učinkovito kreiranje vizualizacija i izvješća.



Slika 15: Prikaz radnog okruženja alata Tableau. (izvor: vlastita izrada)

Kada se otvori alat Tableau na lijevoj strani se može uočiti ploča s podacima („*engl. Dana Pane*“) koji su učitani u ovom slučaju iz Excel tablica koje su spremljene u obradi i analizi podataka. Ploča se sastoji od dvije vrste podataka: **dimenzije** („*engl. Dimensions*“) koje predstavljaju podatke prema kojima se podaci segmentiraju kao na primjer ime države, vrsta goriva i slično, te **mjere** („*engl. Measure*“) koje predstavljaju numeričke varijable koje se mogu agregirati kao na primjer cijena, datum i slično. Odmah pokraj ploče s podacima nalaze se komponente pod nazivom **stranice** („*engl. Pages*“), **filteri** („*engl. Filters*“) i **oznake** („*engl. Marks*“). Komponenta stranice omogućava animaciju podataka na način da se povuče dimenzija na ovu karticu i vizualizacija će se prikazivati stranicu po stranicu za svaku vrijednost te dimenzije. Komponenta filteri služi kako bi se postavili filteri koji omogućuju korisnicima da



istraže podatke u vizualizaciji na temelju određenih uvjeta kao na primjer ako se želi izdvojiti samo određena država ili određeno gorivo. Komponenta oznake omogućava korisnicima da promijene način na koji su podaci predstavljeni u vizualizaciji. Ovdje se može kontrolirati boja, veličina, oblik i tekst određene dimenzije ili mjere.

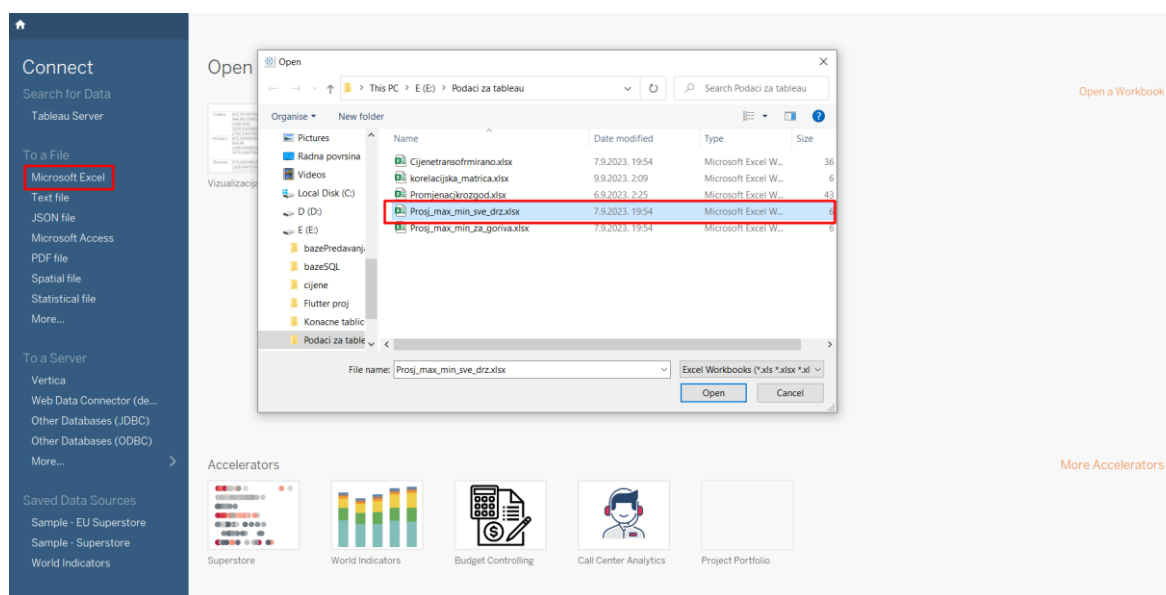
Na gornjoj ploči se može uočiti **alatna traka** koja sadrži opcije poput spremanja i izvoza trenutno radnog lista kao i učitavanje podataka za radni list te objavljivanje vizualizacije na Tableau Server ili Tableau Public koji omogućavaju dijeljenje i suradnju putem interneta sa ostalim korisnicima.

U sredini se nalazi prostor gdje se mogu povlačiti i ispuštati dimenzije i mjere te prema tome izrađivati vlastite vizualizacije. Također u gornjem središnjem prostoru se nalaze prostori za ispuštanje mjera i dimenzija u **stupce** („*engl. Columns*“) ili u **redove** („*engl. Rows*“) ovisno o potrebi i vrsti grafa. Na desnoj strani također postoji i legenda koja se može uključiti da pobliže opisuje prikazani graf te opcija **Show Me** gdje se može odabrati vrsta grafa.

Pri dnu su vidljivi radni listovi koji se mogu kreirati, duplicirati i izvesti slično kao i u Excelu. Za svaki radni list moguće je koristiti drugačiju vrstu podataka jednostavnim odabirom izvora podataka na ploči za podatke.

## 5.2. Vizualizacija prosječne cijene

Prije same izrade grafikona potrebno je učitati podatke koji su prethodno obrađeni, postupak učitavanja podataka vidljiv je na sljedećoj slici:

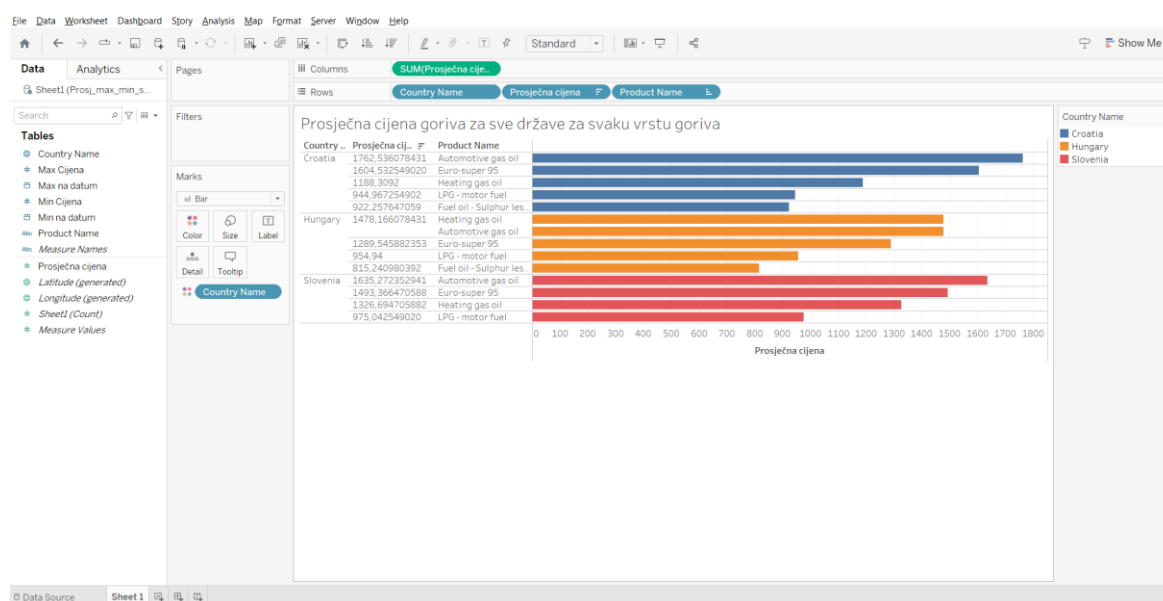


Slika 16: Prikaz učitavanja podataka u alat Tableau. (izvor: vlastita izrada)



Najprije je potrebno izabrati koju vrstu datoteke želimo učitati na način da se otvori početna stranica na ikonicu kućice gdje se nakon odabira otvara početna stranica u alatu Tableau i odabire se željena vrsta datoteke u ovom slučaju je to Excel datoteka. Odabire se prethodno spremljena datoteka koja je obrađena **Pros\_j\_max\_min\_sve.xlsx** te se učitava i spremna je za korištenje u novom kreiranom radnom listu.

Na sljedećoj slici prikazano je radno okruženje izrade grafikona za izračunate prosječnih cijena goriva za sve države:

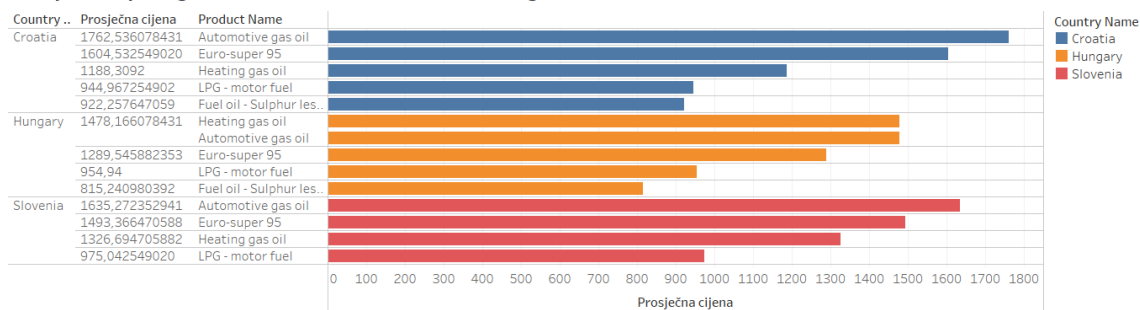


**Slika 17:** Prikaz izrade vizualizacije prosječnih cijena. (izvor:vlastita izrada)

Za ovaj primjer korištena je trakasta vrsta grafikona („engl. *Horizontal Bars*“). Kako bi se ovaj graf napravio najprije je potrebno povući dimenzije „*Country Name*“, „*Prosječna cijena*“ i „*Product Name*“ na traku za redove pa se na taj način prikazuje naziv države, prosječna cijena te naziv goriva. U traci za stupce se povlači mjera „*Prosječna cijena*“ te se tom radnjom dobiva ovakav prikaz grafa, no bez boje koje pobliže opisuju graf stoga je potrebno postaviti dimenziju „*Country Name*“ na oznaku za boju koja će odrediti tri boje koje za države koje se prikazuju te će se stvoriti legenda koja opisuje svaku boju tako da je vidljivije o kojoj se državi radi u grafu. Rezultat grafa se može izvesti odabirom na alatnoj traci **Worksheets>Export>** željeni format u ovom slučaju to slika u PNG formatu.

### 5.3. Interpretacija prosječnih cijena

Prosječna cijena goriva za sve države za svaku vrstu goriva



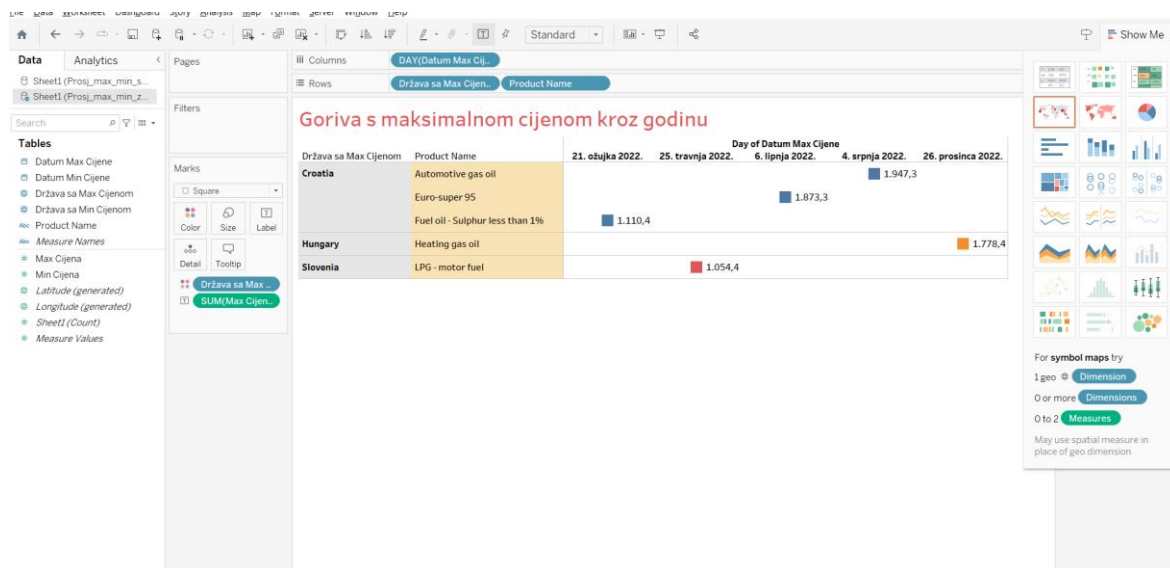
Slika 18: Prikaz vizualizacije prosječnih cijena

Na slici 18 je korišten trakasti tip grafa. U analizu su uključene države Hrvatska, Slovenija i Mađarska koje su označene bojama redom plava, narančasta i crvena. Graf obuhvaća pet različitih vrsta goriva „Automotive gas oil“, „Euro-super 95“, „Heating gas oil“, „LPG – motor fuel“, te „Fuel oil“.

Iz grafa se može zaključiti da je prosječna cijena „Automotive gas oil“ približno podjednako najskuplja u svim državama, izdvojeno najviše u Hrvatskoj. Hrvatska generalno ima najviše prosječne cijene goriva u odnosu na sve tri države dok Mađarska ima prosječno najjeftinije cijene goriva, Slovenija se s druge strane pozicionirala negdje između Hrvatske i Mađarske prema visini prosječnih cijena goriva. Ako se promotri vrsta goriva LPG vidljivo da su cijene vrlo slične u sve tri države što ukazuje na manju varijabilnost cijene za tu vrstu goriva.

### 5.4. Vizualizacija minimalne i maksimalne cijene

U nastavku će biti vizualizirane minimalne i maksimalne cijene goriva kroz godinu. Minimalne i maksimalne cijene odnose se na sva goriva te će biti prikazano na koji datum i koje gorivo je imalo minimalnu ili maksimalnu cijenu te u kojoj je državi to nastupilo. U nastavku na sljedećoj slici biti će opisan postupak izrade vizualizacije:

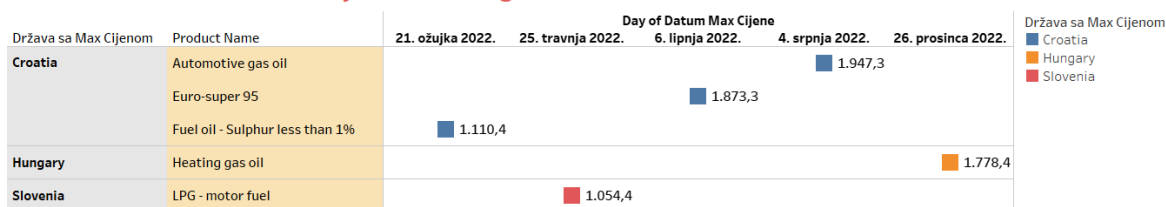


**Slika 19:** Prikaz izrade vizualizacije goriva s maksimalnom cijenom kroz godinu. (izvor: vlastita izrada)

Za izradu ovakvog grafikona korišten je stupčasti grafikon sa kvadratićima. Najprije je potrebno povući dimenziju „Datum Max Cijene“ na traku za stupce te postaviti datum da prikazuje dan, mjesec i godinu. Na traku za redove potrebno je povući dimenzije „Država sa Max Cijenom“ i „Product Name“ za vrstu goriva, time se ostvaruje prikaz gdje su nazivi država i nazivi vrsta goriva u redovima dok se datumi prikazuju po stupcima za svaku minimalnu ili maksimalnu cijenu koja je nastala na taj datum. U oznake je potrebno staviti dimenziju sa državom na komponentu za boje kako bi se države razlikovale po boji te je potrebno staviti mjeru „Max Cijena“. Isti postupak je i za vizualizaciju minimalne cijene.

## 5.5. Interpretacija minimalne i maksimalne cijene

### Goriva s maksimalnom cijenom kroz godinu



Sum of Max Cijena broken down by Datum Max Cijene Day vs. Država sa Max Cijenom and Product Name. Color shows details about Država sa Max Cijenom. The marks are labeled by sum of Max Cijena.

**Slika 20:** Prikaz vizualizacije goriva s maksimalnom cijenom kroz godinu. (izvor: vlastita izrada)

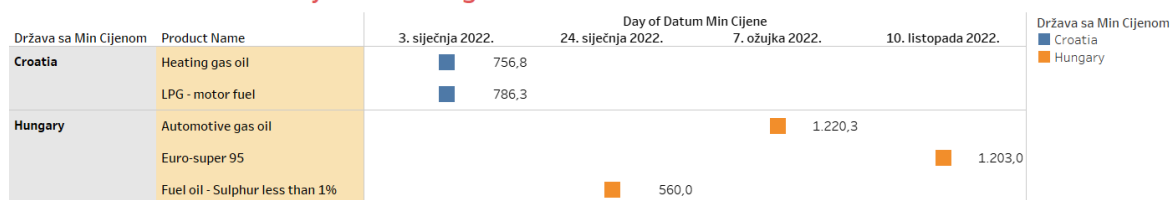
Na prethodnoj slici je stupčasti grafikon koji prikazuje maksimalne cijene pet različitih vrsta goriva („Automotive Gas Oil“, „Euro-Super 95“, „Fuel Oil“, „Heating Gas Oil“ i „LPG Motor Fuel“) u tri različite države (Hrvatska, Mađarska i Slovenija) za godinu 2022.

Uočljivo je da „Automotive Gas Oil“ i „Euro-Super 95“ imaju najviše cijene u sve tri države, s posebnim naglaskom na Hrvatsku gdje su cijene znatno više u usporedbi s Mađarskom i Slovenijom. „Fuel Oil“ i „Heating Gas Oil“ imaju slične cijene u sve tri države, ali su nešto više u Hrvatskoj. „LPG Motor Fuel“ ima najniže cijene u svim tri države, ali je najskuplji u Sloveniji.

Boje na grafikonu koriste se za razlikovanje između različitih država, dok su različite vrste goriva predstavljene odvojenim stupcima. Ovo omogućuje brzu vizualnu usporedbu i identifikaciju trendova ili odstupanja. Ukupno gledano, grafikon pruža jasnu sliku o tome kako se cijene različitih vrsta goriva razlikuju između država. Posebno je zanimljivo što Hrvatska ima generalno više cijene za većinu vrsta goriva u usporedbi s ostalim dvjema državama.

Na sljedećoj slici prikazan je grafikon sa minimalnim cijenama:

#### Goriva s minimalnom cijenom kroz godinu



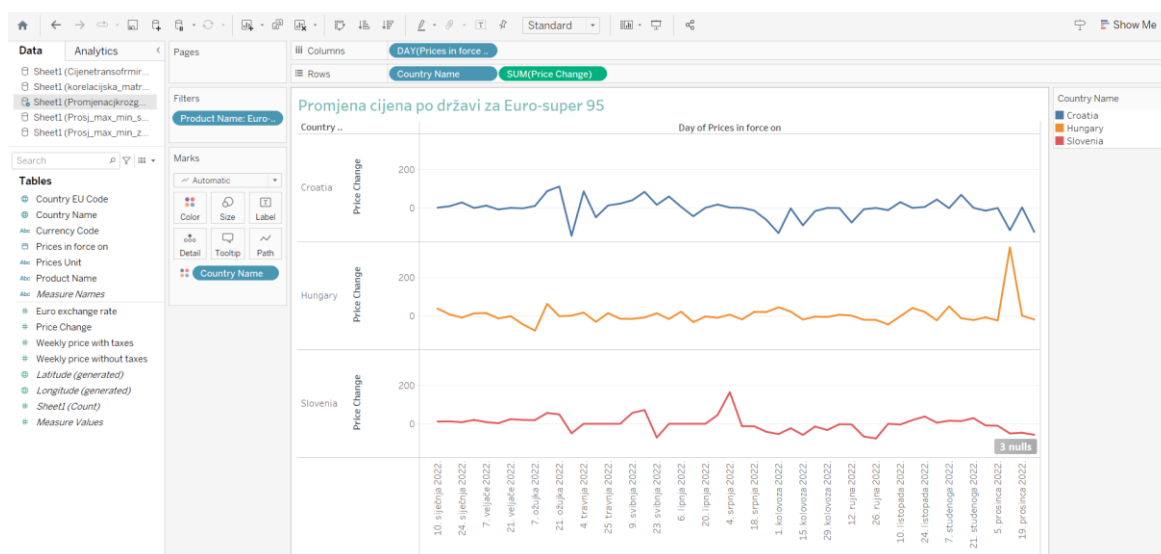
Sum of Min Cijena broken down by Datum Min Cijene Day vs. Država sa Min Cijenom and Product Name. Color shows details about Država sa Min Cijenom. The marks are labeled by sum of Min Cijena.

**Slika 21:** Prikaz vizualizacije goriva s minimalnom cijenom kroz godinu. (izvor: vlastita izrada)

U sve tri države, „Automotive Gas Oil“ i „Euro-Super 95“ imaju slične minimalne cijene, s iznimkom Slovenije gdje je „Automotive Gas Oil“ znatno skuplji. „LPG Motor Fuel“ ima najniže cijene u svim tri države, ali najniže cijene su zabilježene u Hrvatskoj. „Fuel Oil“ i „Heating Gas Oil“ također imaju slične minimalne cijene u svim državama, s nešto nižim cijenama u Mađarskoj. Hrvatska i Mađarska imaju slične minimalne cijene za većinu vrsta goriva, dok Slovenija ima nešto više cijene za „Automotive Gas Oil“.

## 5.6. Promjena cijena po državi

U ovoj vizualizaciji koristiti će se linijski grafikon s ciljem prikaza promjena cijena za svaki tjedan te za svaku vrstu goriva, promjena je izračunata na način da se cijena u tekućem tjednu oduzima od cijene prethodnog tjedna. Promjene u prvom tjednu nije moguće izračunati jer nema prethodne cijene sa kojom bi se mogla izračunati promjena cijene stoga će one biti NULL (prazne vrijednosti). Na sljedećoj slici prikazan je grafikon u alatu Tableau:

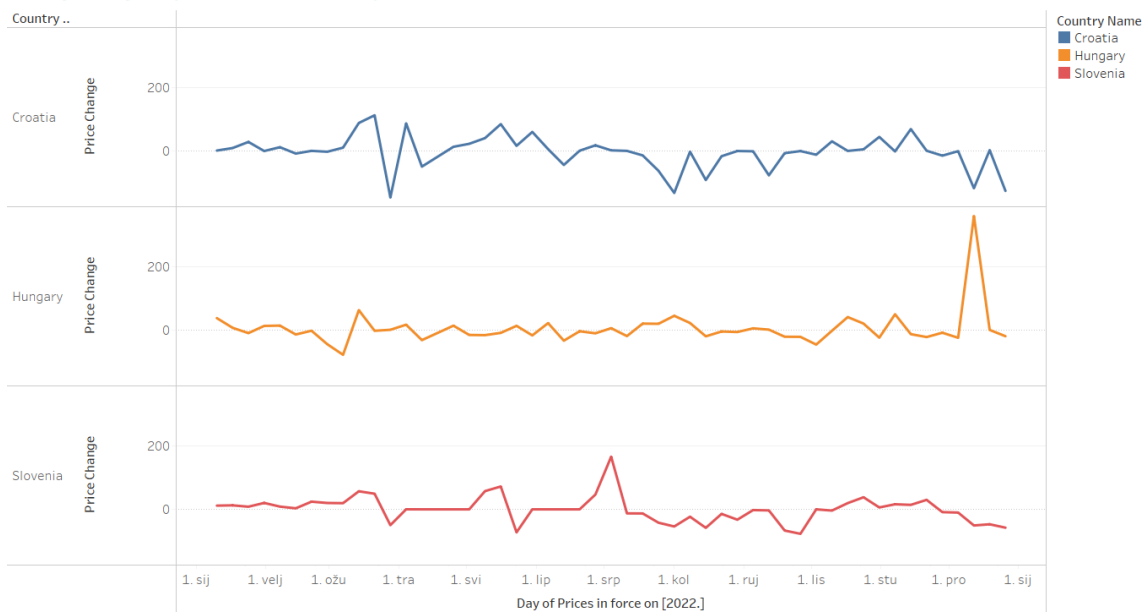


Slika 22: Prikaz postupka izrade vizualizacije promjene cijena po državi. (izvor: vlastita izrada)

Za izradu ovog linijskog grafikona potrebno je povući dimenziju „*Prices in force on*“ koja predstavlja vremensku varijablu na horizontalnu os. Na vertikalnu os potrebno je povući dimenziju „*Country Name*“ za prikaz naziva države te mjeru „*Price change*“ za promjenu cijene. Također dimenziju za naziv države potrebno je povući u karticu oznaka u boje kako bi se linije država razlikovale po boji. Na posljertku s obzirom da se gleda promjena cijene goriva za svaku vrstu goriva posebno u ovom slučaju se promatra „*Euro Super 95*“ stoga je u filtere potrebno filtrirati goriva prema vrsti goriva i odabrati onu vrstu goriva koja se promatra. Isti postupak izrade je i za ostale grafikone za promjenu cijena goriva.

## 5.7. Interpretacija promjene cijene prema vrsti goriva

Promjena cijena po državi za Euro-super 95

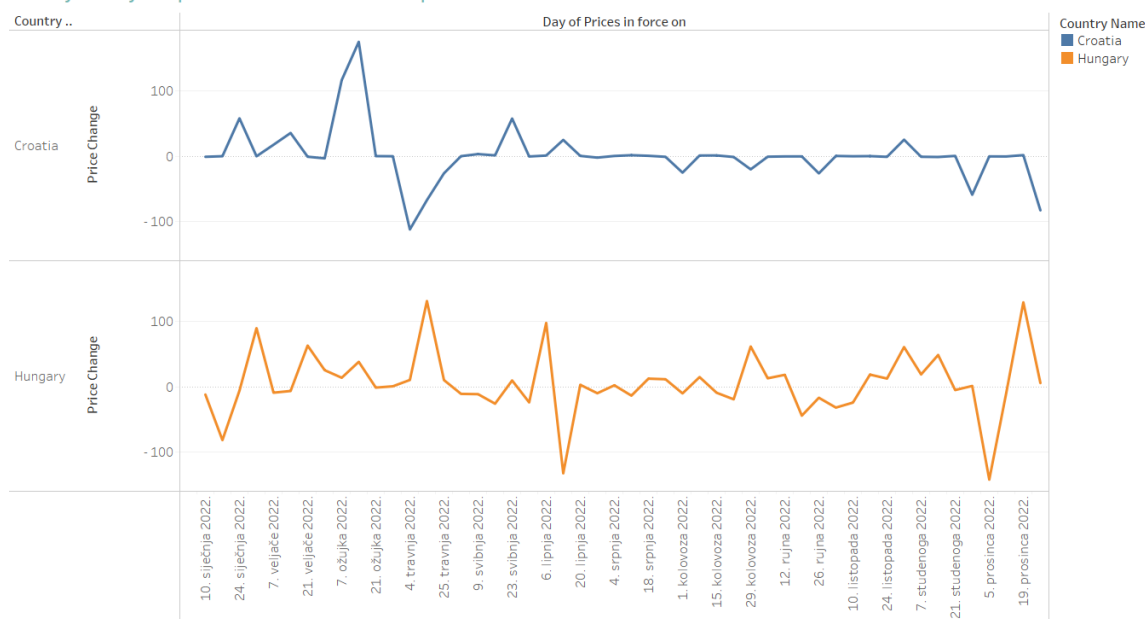


The trend of sum of Price Change for Prices in force on Day broken down by Country Name. Color shows details about Country Name. The data is filtered on Product Name, which keeps Euro-super 95.

**Slika 23:** Prikaz vizualizacije promjene cijene prema vrsti goriva 1.primjer. (izvor: vlastita izrada)

Na slici je vidljiv grafikon koji prikazuje kako se cijene goriva vrste „Euro-super 95“ mijenjaju kroz vremenski period u tri različite države: Hrvatska, Mađarska i Slovenija. Na grafikonu je korištena vremenska serija kako bi se prikazale dinamike promjena cijena. Iz grafikona se može zaključiti da Hrvatska pokazuje znatne fluktuacije u promjenama cijena, s nekoliko uspona i padova, slično je i sa Slovenijom ali sa manje izraženim usponima i padovima. U mađarskoj su fluktuacije cijena najstabilnije s manje izraženim promjenama.

#### Promjena cijena po državi za Fuel oil - Sulfur less than 1%

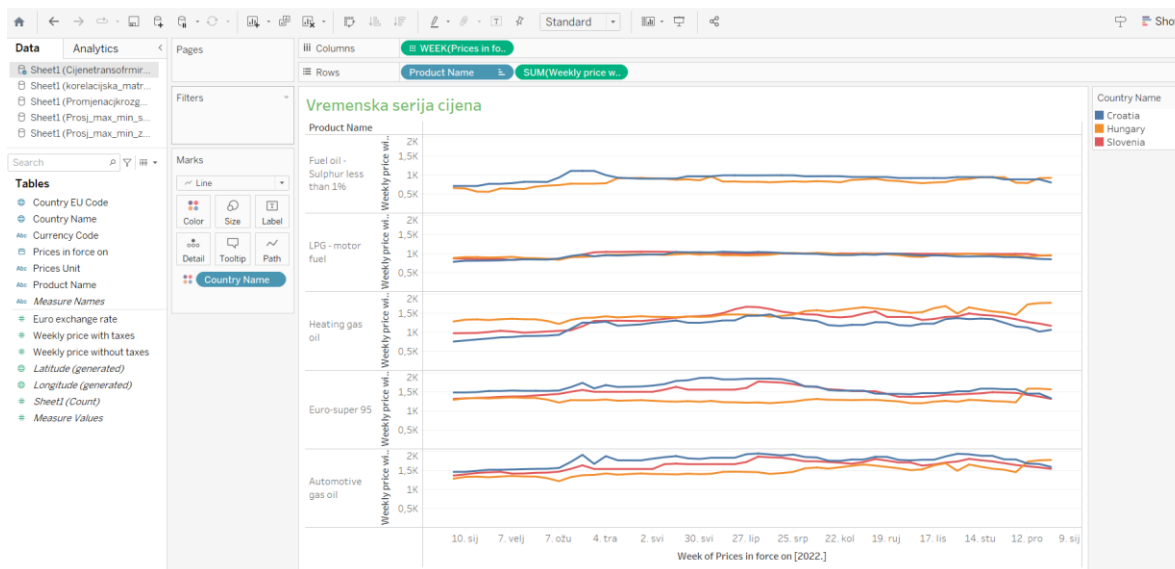


**Slika 24:** Prikaz vizualizacije promjene cijene prema vrsti goriva 2.primjer. (izvor: vlastita izrada)

Na slici je vidljiv grafikon koji prikazuje kako se cijene goriva vrste „Fuel oil – Sulfur less than 1%“ mijenjaju kroz vremenski period u dvije različite države: Hrvatska i Mađarska jer u Sloveniji ovakva vrsta goriva se ne prodaje. Iz grafa se može zaključiti da su u hrvatskoj promjene cijena relativno male i konzistentne osim na početku godine gdje su vidljivi malo veći uspon i pad, no u ostatku se može zaključiti da su promjene cijena za ovu vrstu goriva relativno stabilne. S druge strane u mađarskoj graf pokazuje puno veću nesigurnost i fluktuaciju cijena jer su kroz godinu vidljivi nekoliko značajnih uspona i padova što ukazuje na veću nestabilnost cijena i moguću osjetljivost na vanjske faktore u toj državi za tu vrstu goriva. **Prikaz ostalih vizualizacija za promjenu cijena goriva nalaze se u prilogu Vizualizacija.twb na radnom listu promjena cijena po državi.**

## 5.8. Vremenska serija cijena

Izrada vizualizacije vremenske serija cijena za goriva daje bolji pregled cijena kroz godinu linijskim grafikonom. Na sljedećoj slici vidljiv je prikaz grafikona u alatu Tableau:



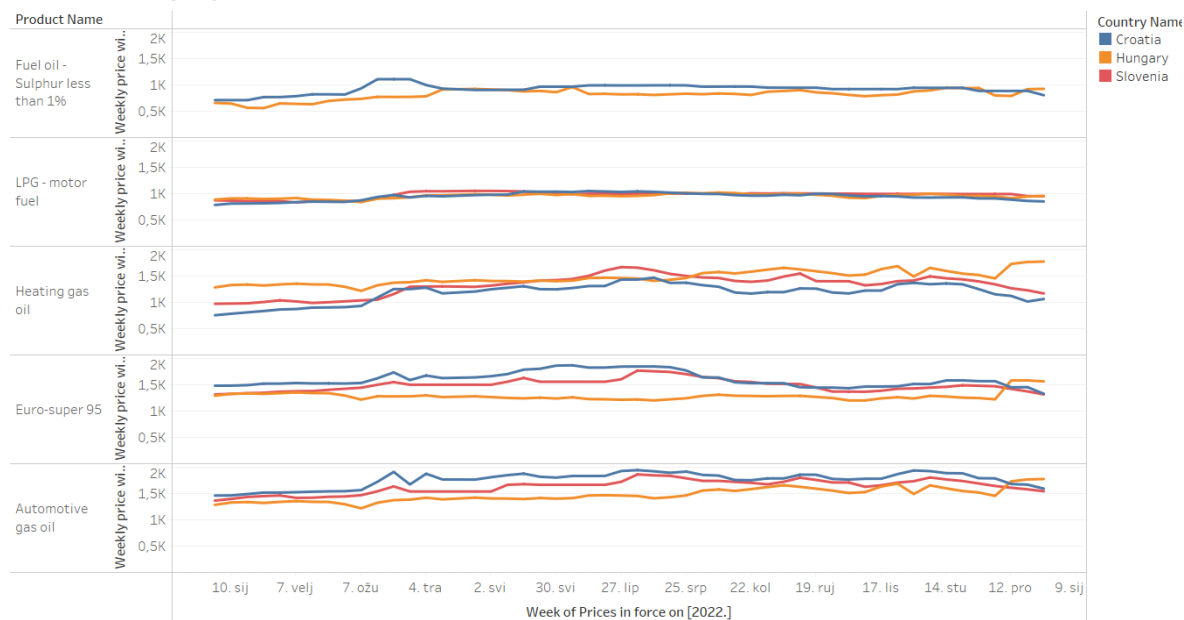
Slika 25: Prikaz postupka izrade vremenskih serija cijena. (izvor: vlastita izrada)

Za izradu ovog linijskog grafikona potrebno je povući dimenziju datuma na horizontalnu os kako bi dobili datume nastanka cijena kroz cijelu godinu. Dimenziju za nazive goriva potrebno je povući na vertikalnu os kao i mjeru tjednih cijena, mjeru je potrebno postaviti na opciju dvojne osi („engl. *Dual axis*“) kako bi vremenski tijek cijena za svaku vrstu goriva bio prikazan na istom grafikonu. Linije se razlikuju po boji na način da se postavi filter boja na dimenziju naziva svake vrste goriva. Ovim postupkom se ostvaruje jednostavan linijski prikaz cijena kroz godinu te je lakše uočiti razlike za svaku državu.



## 5.9. Interpretacija vremenske serije cijena

### Vremenska serija cijena



Slika 26: Prikaz vizualizacije vremenske serije cijena. (izvor: vlastita izrada)

Na linijskom grafikonu vidljive su cijene svih vrsta goriva u sve tri države te se razlikuju prema boji. Hrvatska je predstavljena plavom bojom, Mađarska narančastom i Slovenija Crvenom. Treba uzeti u obzir da „Fuel oil“ lož ulje nema cijenu u Sloveniji.

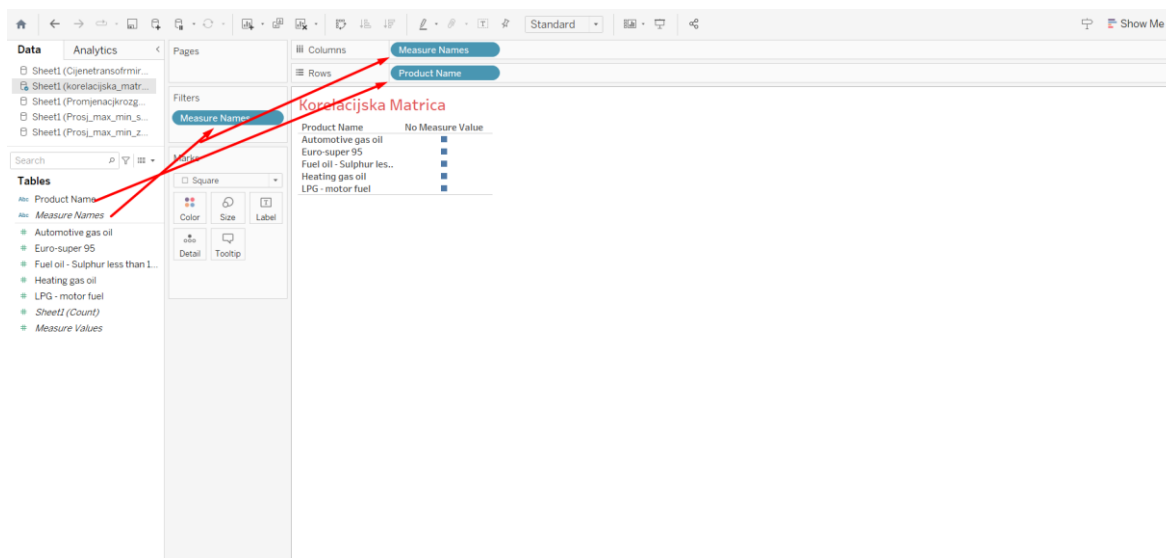
Iz grafa je zapaženo da za vrstu goriva lož ulja Hrvatska i Slovenija prikazuju poprilično jednake stabilne cijene, Hrvatska ima blago povišene cijene u ožujku i travnju. Cijene kroz godinu vrste goriva „LPG“ u Hrvatskoj, Sloveniji i Mađarskoj su najstabilnije i jednake kroz sve tri države, moguće je da je razlog tomu što je stabilna ponuda i potražnja za ovom vrstom goriva stoga su cijene manje podložne promjenama. Cijene vrste goriva „Heating gas oil“ u Hrvatskoj i Sloveniji su približno jednake dok su u Mađarskoj vidljive veće oscilacije vjerojatno zbog sezonskih potreba jer je vidljivo najveći skok cijene u prosincu. Cijene vrste goriva „Euro super 95“ u Hrvatskoj i Sloveniji blago rastu sve do srpnja gdje cijene blago opadaju do kraja godine. U Mađarskoj cijena je stabilna kroz cijelu godinu dok je zabilježen nagli skok na kraju godine. Cijene za vrstu goriva „Automotive gas oil“ u Hrvatskoj su relativno stabilne na početku godine, s blagim porastom u sredini. Međutim, prema kraju godine vidljiv je pad cijena. Ova promjena može ukazivati na sezonske faktore ili promjene u ponudi i potražnji. U Mađarskoj cijene se kreću blago uzlazno tijekom većeg dijela godine, s naglim porastom prema kraju. Ovo može ukazivati na povećanu potražnju ili smanjenje ponude u toj vremenskoj fazi. U Sloveniji cijene su također relativno stabilne, ali s manjim varijacijama u odnosu na druge dvije države.

## 5.10. Korelacijska matrica

Korelacijska matrica prikazuje linearnu povezanost između različitih vrsta goriva na temelju njihovih tjednih cijena s porezom. Vrijednosti u matrici kreću se od -1 do 1:

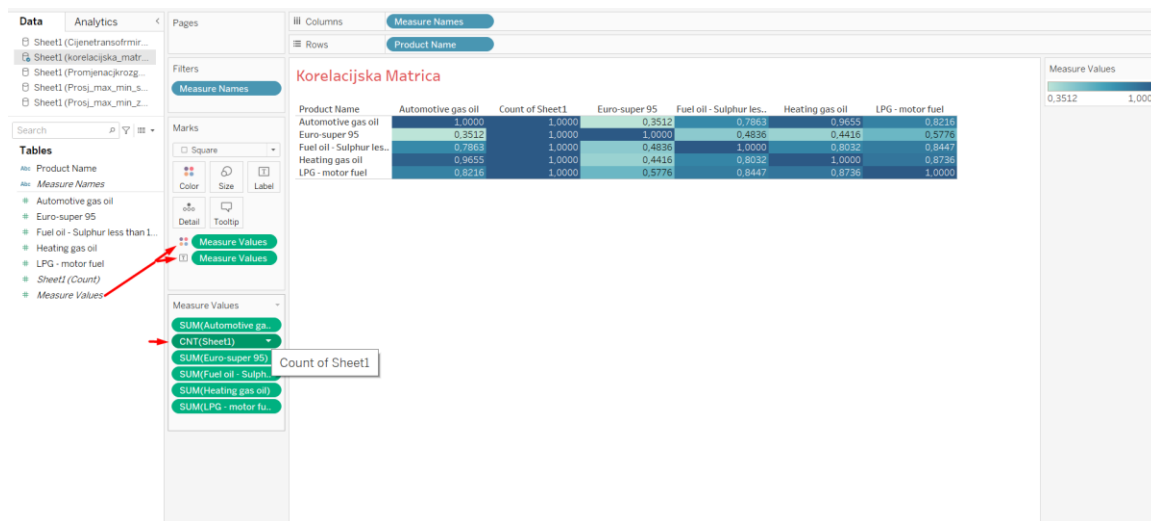
- Vrijednost 1 označava potpunu pozitivnu korelaciju: ako cijena jednog goriva raste, cijena drugog goriva također raste.
- Vrijednost -1 označava potpunu negativnu korelaciju: ako cijena jednog goriva raste, cijena drugog goriva opada.
- Vrijednost blizu 0 označava nisku ili nikakvu korelaciju između cijena dviju vrsta goriva.

Za prikaz korelacijske matrice koristi se toplinska tablica („*engl. Heat map*“). S obzirom da je postupak izrade ovakve vizualizacije malo zahtjevniji opis izrade će biti detaljniji u nastavku.



Slika 27: Prikaz postupka izrade vizualizacije korelacijske matrice 1.dio. (izvor: vlastita izrada)

Na slici su prikazane strelice koje prikazuju dimenzije koje je potrebno koristiti. Dimenzija „*Measure Names*“ predstavlja nazive vrijednosti goriva te ju je potrebno povući na horizontalnu os te na filteru označiti sva goriva koja se prikazuju u grafikonu te postaviti oznaku („*engl. Marks*“) na kvadrat („*engl. square*“), nazivi vrste goriva se povlače na vertikalnu os.



Slika 28: Prikaz postupka izrade vizualizacije korelacijske matrice 2.dio. (izvor: vlastita izrada)

Na prethodnoj slici se može vidjeti da je potrebno staviti mjere „Measure values“ na karticu oznaka i dodati im boju i tekst kako bi se dobio toplinski prikaz korelacijske matrice. Treba napomenuti da je potrebno obrisati mjeru „CNT“ jer ona nije potrebna u ovom prikazu grafikona.

## 5.11. Interpretacija korelacijske matrice

### Korelacijska Matrica

Product Name	Automotive gas oil	Euro-super 95	Fuel oil - Sulphur les..	Heating gas oil	LPG - motor fuel
Automotive gas oil	1,0000	0,3512	0,7863	0,9655	0,8216
Euro-super 95	0,3512	1,0000	0,4836	0,4416	0,5776
Fuel oil - Sulphur les..	0,7863	0,4836	1,0000	0,8032	0,8447
Heating gas oil	0,9655	0,4416	0,8032	1,0000	0,8736
LPG - motor fuel	0,8216	0,5776	0,8447	0,8736	1,0000

Automotive gas oil, Euro-super 95, Fuel oil - Sulphur less than 1%, Heating gas oil and LPG - motor fuel broken down by Product Name. Color shows Automotive gas oil, Euro-super 95, Fuel oil - Sulphur less than 1%, Heating gas oil and LPG - motor fuel. The marks are labeled by Automotive gas oil, Euro-super 95, Fuel oil - Sulphur less than 1%, Heating gas oil and LPG - motor fuel.

Slika 29: Prikaz vizualizacije korelacijske matrice. (izvor: vlastita izrada)

Na slici je prikazana vizualizacija korelacijske matrice toplinskom tablicom zapažene vrijednosti su sljedeće:

„Automotive gas oil“:

- Najveća korelacija: S "Heating gas oil" (0.9655) – vrlo visoka pozitivna korelacija ukazuje na vrlo slične trendove u cijenama.

- Najmanja korelacija: S "*Euro-super 95*" (0.3512) – niska pozitivna korelacija ukazuje na slabu povezanost u kretanju cijena.

„*Euro Super 95*“:

- Najveća korelacija: S "*LPG - motor fuel*" (0.5776) – srednja pozitivna korelacija ukazuje na umjerenu povezanost u kretanju cijena.
- Najmanja korelacija: S "*Heating gas oil*" (0.4416) – niska pozitivna korelacija ukazuje na slabu povezanost u kretanju cijena.

„*Fuel oil - Sulphur less than 1%*“:

- Najveća korelacija: S "*LPG - motor fuel*" (0.8447) – visoka pozitivna korelacija ukazuje na slične trendove u cijenama.
- Najmanja korelacija: S "*Euro-super 95*" (0.4836) – srednja pozitivna korelacija ukazuje na umjerenu povezanost u kretanju cijena.

„*Heating gas oil*“:

- Najveća korelacija: S "*Automotive gas oil*" (0.9655) – vrlo visoka pozitivna korelacija ukazuje na vrlo slične trendove u cijenama.
- Najmanja korelacija: S "*Euro-super 95*" (0.4416) – niska pozitivna korelacija ukazuje na slabu povezanost u kretanju cijena.

„*LPG - motor fuel*“:

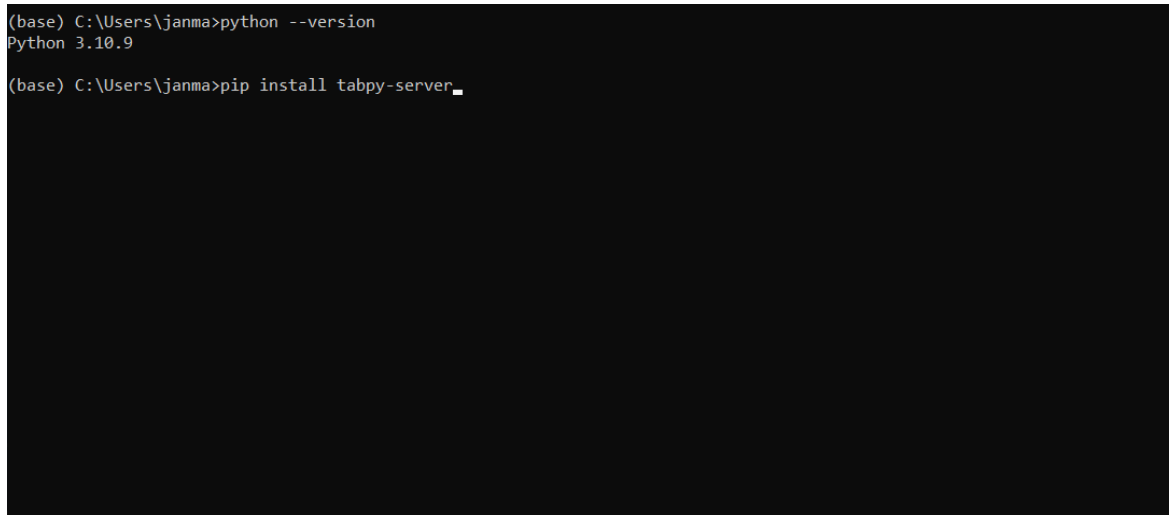
- Najveća korelacija: S "*Heating gas oil*" (0.8736) – visoka pozitivna korelacija ukazuje na slične trendove u cijenama.
- Najmanja korelacija: S "*Euro-super 95*" (0.5776) – srednja pozitivna korelacija ukazuje na umjerenu povezanost u kretanju cijena.

Iz toplinske tablice može se zaključiti da najveću prosječnu pozitivnu korelaciju ima „*Automotive gas oil*“ dok gorivo „*Euro-super 95*“ ima najmanju prosječnu pozitivnu korelaciju. Visoka pozitivna korelacija između dvije vrste goriva ukazuje da se njihove cijene obično kreću u istom smjeru. Ako jedno gorivo doživi porast cijene, vrlo je vjerojatno da će i drugo imati porast. Srednja korelacija ukazuje na umjerenu povezanost, dok niska korelacija ukazuje na to da su promjene cijena među vrstama goriva manje povezane.

Ovakav tip analize može biti koristan za razumijevanje kako se tržišni faktori odražavaju na cijene različitih vrsta goriva, što može biti korisno za donošenje informiranih odluka o kupnji ili prodaji.

## 5.12. Razlika cijene i poreza

U ovom primjeru promatrati će se razlika između cijene goriva sa porezom i cijene goriva bez poreza. Prilikom ovog izračuna korišten je Python programski jezik unutar alata Tableau pomoću TabPy-a. TabPy je vanjski dodatak u alatu Tableau koji omogućuje izvršavanje Python kodova iz Tableau-a. TabPy je osmišljen kako bi se proširile mogućnosti Tableau-a u domeni analitičkih algoritama i modela koje Python podržava. Kad korisnik iz Tableau-a pozove Python skriptu, upit se šalje na TabPy server, koji zatim izvršava Python kod i vraća rezultate natrag u Tableau. Kako bi uspjeli izvršavati Python kod unutar Tableau potrebno je instalirati dodatak TabPy u ovom slučaju instaliran je preko naredbenog retka naredbom **pip install tabpy-server** kao što je prikazano na sljedećoj slici:

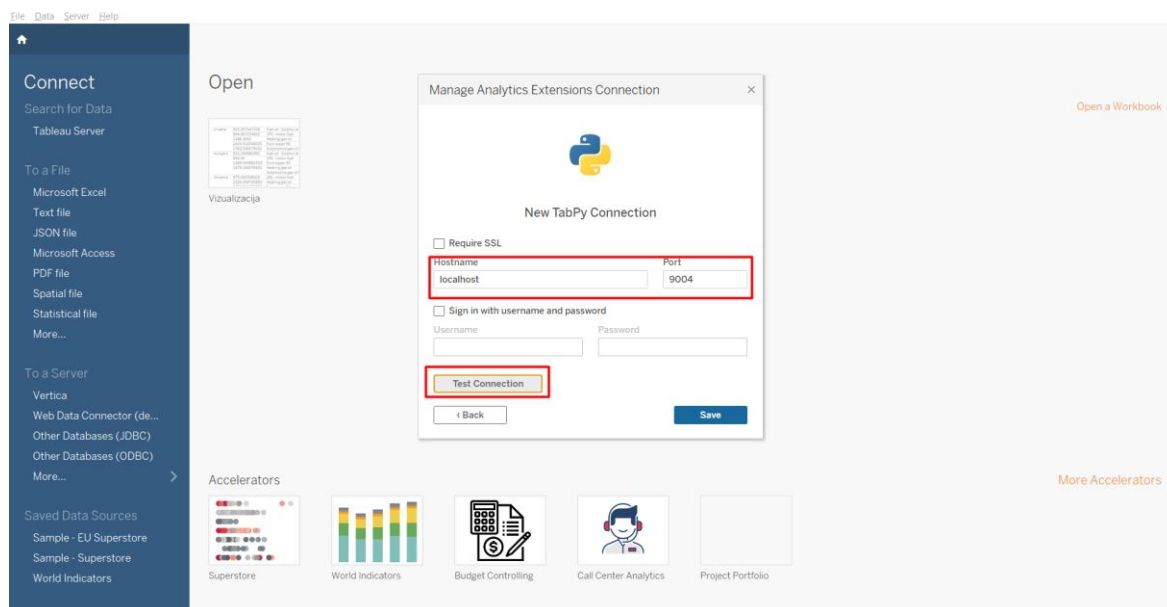


```
(base) C:\Users\janma>python --version
Python 3.10.9

(base) C:\Users\janma>pip install tabpy-server_
```

**Slika 30:** Prikaz naredbe za instalaciju TabPy dodatka (izvor: vlastita izrada)

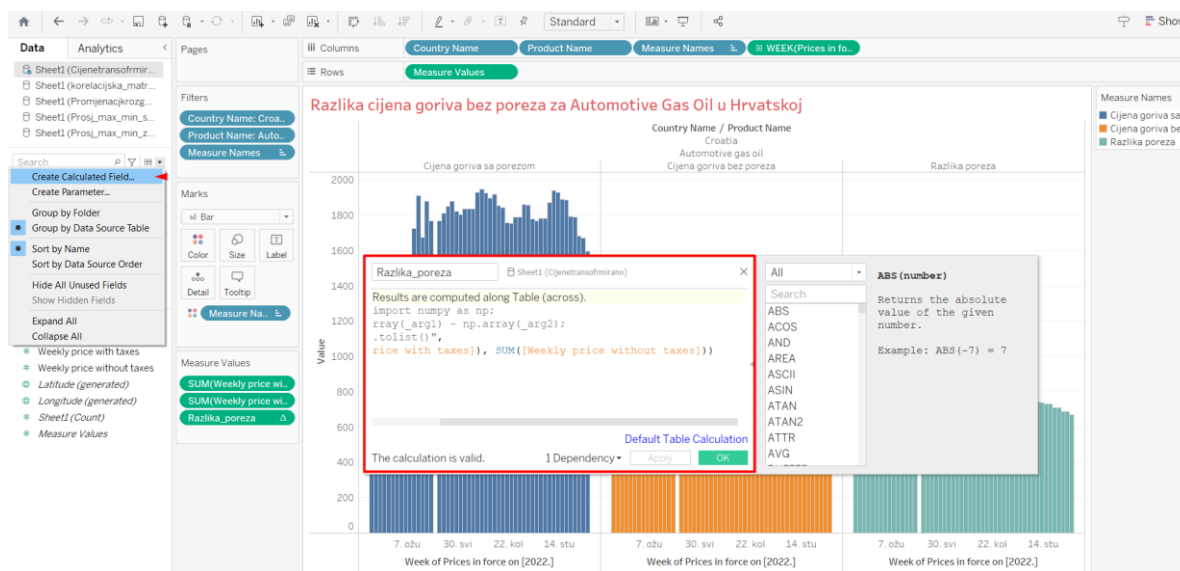
Nakon instalacije potrebno je pokrenuti TabPy server pomoću naredbe `tabpy`. TabPy server je potrebno pokrenuti prije pokretanja alata Tableau i povezati se na server pomoću Tableau kao na slici 31:



**Slika 31:** Prikaz povezivanja alata Tableau na TabPy server (izvor: vlastita izrada)

Postupak povezivanja se radi na način da se odabere opcija **Manage analytics Extensions Connection** gdje se pod naziv hosta upisuje localhost te port koji piše u naredbenom retku u ovom slučaju je to 9004.

U nastavku izrada vizualizacije daje bolji pregled u razlikama cijene goriva sa porezom i bez poreza. Na sljedećoj slici prikazan je način izrade vizualizacije za „Automotive Gas Oil“ u Hrvatskoj:

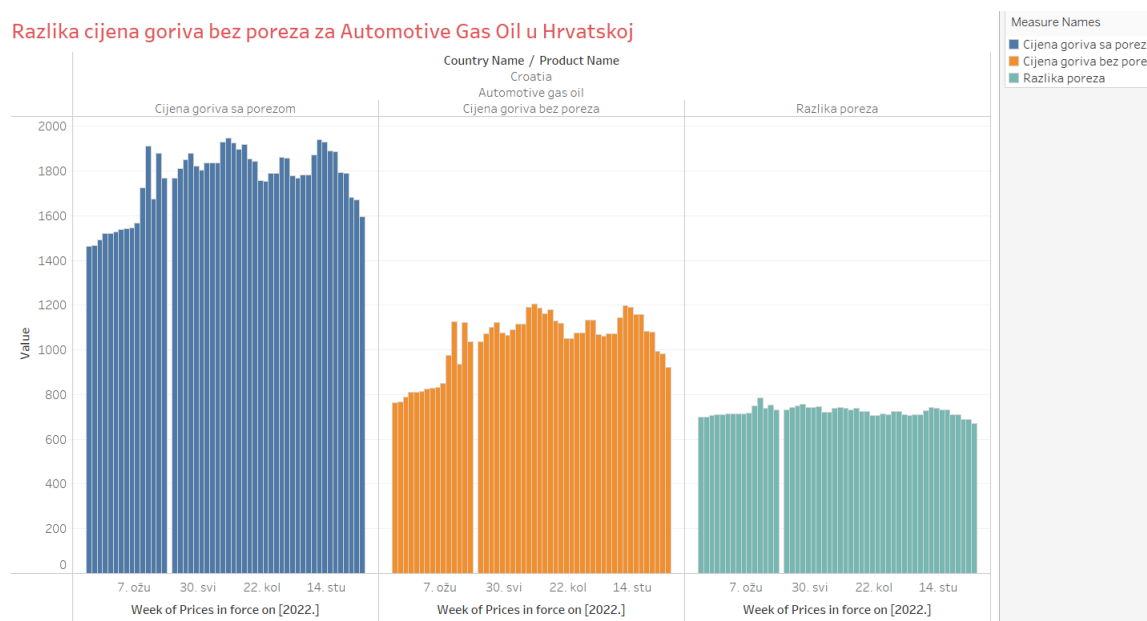


**Slika 32:** Prikaz postupka izrade vizualizacije za razliku cijene i poreza goriva (izvor: vlastita izrada)

Za izradu ove vizualizacije najprije je potrebno dodati novu mjeru koja će se koristiti za izračun razlike cijene goriva. U ovom slučaju dodana je mjera „Razlika poreza“ pomoću opcije („*engl. Create calculated field*“) gdje je moguće upisati Python kod koji će se izvršavati nad proslijeđenim argumentima. U ovom slučaju korištena je SCRIPT\_REAL funkcija koja omogućava izvršavanje Python koda u alatu Tableau. Za izračun razlike korištena je biblioteka **NumPy** koja služi za numeričke izračune. U varijablu **result** se sprema razlika nizova kojima su proslijeđeni argumenti „*Weekly price with taxes*“ i „*Weekly price without taxes*“. Ova mjera se sprema te se koristi u vizualizaciji. Dimenzije „*Country Name*“, „*Product Name*“ i „*Measure Names*“ te mjeru za datum potrebno je povući na horizontalnu os. Mjera „*Measure values*“ se povlači na vertikalnu os te se u njoj nalazi i mjera razlika poreza pomoću koje se vrši izračun.

### 5.13. Interpretacija razlike cijene goriva i poreza

Na sljedećoj slici prikazana je vizualizacija razlike cijene goriva i poreza za „*Automotive Gas Oil*“ u Hrvatskoj:



**Slika 33:** Prikaz vizualizacije razlike cijene goriva i poreza. (izvor: vlastita izrada)

Plavom bojom označene su cijene sa porezom. Ove cijene su vidno više od cijena bez poreza tijekom cijele godine, što je očekivano s obzirom na porezne stope. Narančastom bojom označene su cijene bez poreza. One su konstantno niže od cijena sa porezom, ali se također mijenjaju tijekom godine, što može ukazivati na fluktuacije u nabavnoj cijeni. Svijetlozelenom bojom prikazana je razlika između cijena sa porezom i bez poreza. Iz grafa je vidljivo da su promjene uglavnom stabilne sa minimalnim promjenama što znači da razlika cijene goriva sa porezom i bez poreza kroz godinu ostaje poprilično konstantne.

## 6. Zaključak

Tema ovog završnog rada bila je analiza cijena goriva u tri različite države – Hrvatskoj, Sloveniji i Mađarskoj – koristeći Python u Jupyter Notebooku i vizualizaciju podataka u Tableau alatu. Cilj rada bio je proučiti dinamiku i korelacije u cijenama različitih vrsta goriva kroz razdoblje jedne godine, te identificirati trendove i oscilacije koje bi mogle biti korisne u poslovnom odlučivanju. Tijekom analize, korištene su različite statističke metode, uključujući izračunavanje prosječnih, minimalnih i maksimalnih cijena, vremenske serije i korelacijske matrice. Takvom analizom otkriveno je da postoje značajne varijacije u cijenama između različitih vrsta goriva i među različitim državama. Tableau je poslužio kao moćan alat za vizualizaciju ovih podataka. Njegova sposobnost da transformira složene podatke u intuitivne vizualizacije omogućava dublje razumijevanje podataka. Programski jezik Python je omogućio jednostavno učitavanje i manipulaciju velikog skupa podataka te čišćenje i naposljetku izvedbu samih statističkih izračuna. U konačnici, ovaj završni rad služi kao korak prema dubljem razumijevanju dinamike cijena goriva u različitim geografskim i ekonomskim kontekstima. Rezultati mogu biti korisni ne samo za potrošače i distributere goriva, već i za donositelje odluka na različitim razinama – od lokalne do nacionalne što je dokaz da svakom većem poduzeću poslovna inteligencija može biti od velikog značaja u poslovnom odlučivanju.



## Popis literature

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36(4).

Gartner. (2019). Magic Quadrant for Analytics and Business Intelligence Platforms.

Europska komisija. (2022.). Weekly Oil Bulletin. Energy. Preuzeto sa [https://energy.ec.europa.eu/data-and-analysis/weekly-oil-bulletin\\_en](https://energy.ec.europa.eu/data-and-analysis/weekly-oil-bulletin_en)

Murray, D. G. (2016). Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software. Wiley.

Chabot, C., & Selipsky, A. (2019). Tableau For Dummies. Hoboken, NJ: John Wiley & Sons.

Murray, D. (2020). Learning Tableau 2020: Data Visualization and Business Intelligence, Fourth Edition. Birmingham, UK: Packt Publishing.

Pandas Development Team. Pandas Documentation. Preuzeto 5.svibnja.2023. sa <https://pandas.pydata.org/docs>

GeeksforGeeks. Pandas Practice Exercises, Questions and Solutions. Preuzeto 1. ožujka 2023. sa [GeeksforGeeks. \(n.d.\). Pandas Practice Exercises, Questions and Solutions. Preuzeto 1. ožujka 2023., iz https://www.geeksforgeeks.org/pandas-practice-exercises-questions-and-solutions/?source=post\\_page-----b5453cdfc69f-----](https://www.geeksforgeeks.org/pandas-practice-exercises-questions-and-solutions/?source=post_page-----b5453cdfc69f-----)

Brydon, M. J. Multiple Regression in Python. Preuzeto 15. travnja 2023. sa [https://www.sfu.ca/~mjbrydon/tutorials/BAinPy/10\\_multiple\\_regression.html](https://www.sfu.ca/~mjbrydon/tutorials/BAinPy/10_multiple_regression.html)

Kaggle. (2023.). Pandas. Preuzeto 17. travnja 2023. sa <https://www.kaggle.com/learn/pandas>

Cohen, J., Cohen P., West, S.G., & Aiken, L.S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, NJ: Lawrence Erlbaum Associates.

## Popis slika

<b>Slika 1:</b> Prikaz tablice s podacima (izvor: vlastita izrada) .....	5
<b>Slika 2:</b> Prikaz radnog okruženja alata Jupyter Notebook (izvor: vlastita izrada) .....	8
<b>Slika 3:</b> Prikaz učitavanja podataka i biblioteke Pandas (izvor: vlastita izrada) .....	10
<b>Slika 4:</b> Prikaz pretvorbe tipa podataka (izvor: vlastita izrada) .....	11
<b>Slika 5:</b> Prikaz zamjene mjernih jedinica. (izvor: vlastita izrada) .....	11
<b>Slika 6:</b> Prikaz DataFrame nakon promjene. (izvor: vlastita izrada) .....	11
<b>Slika 7:</b> Prikaz programskog koda za izračun prosječne, minimalne i maksimalne cijene. (izvor: vlastita izrada) .....	12
<b>Slika 8:</b> Prikaz DataFrame-a nakon izračuna. (izvor: vlastita izrada) .....	13
<b>Slika 9:</b> Prikaz koda za izračun minimalne i maksimalne cijene prema vrsti goriva. (izvor: vlastita izrada) .....	13
<b>Slika 10:</b> Prikaz DataFrame-a sa izračunima maksimalne i minimalne cijene (izvor: vlastita izrada) .....	14
<b>Slika 11:</b> Prikaz kod za izračun promjene cijena kroz godinu. (izvor: vlastita izrada) .....	14
<b>Slika 12:</b> Prikaz DataFrame-a sa promjenama cijena kroz godinu. (izvor: vlastita izrada) .....	15
<b>Slika 13:</b> Prikaz koda za izračun korelacijske matrice. (izvor: vlastita izrada) .....	15
<b>Slika 14:</b> Prikaz DataFrame-a korelacijske matrice. (izvor: vlastita izrada) .....	16
<b>Slika 15:</b> Prikaz radnog okruženja alata Tableau. (izvor: vlastita izrada) .....	17
<b>Slika 16:</b> Prikaz učitavanja podataka u alat Tableau. (izvor: vlastita izrada) .....	18
<b>Slika 17:</b> Prikaz izrade vizualizacije prosječnih cijena. (izvor: vlastita izrada) .....	19
<b>Slika 18:</b> Prikaz vizualizacije prosječnih cijena .....	20
<b>Slika 19:</b> Prikaz izrade vizualizacije goriva s maksimalnom cijenom kroz godinu. (izvor: vlastita izrada) .....	21
<b>Slika 20:</b> Prikaz vizualizacije goriva s maksimalnom cijenom kroz godinu. (izvor: vlastita izrada) .....	21
<b>Slika 21:</b> Prikaz vizualizacije goriva s minimalnom cijenom kroz godinu. (izvor: vlastita izrada) .....	22
<b>Slika 22:</b> Prikaz postupka izrade vizualizacije promjene cijena po državi. (izvor: vlastita izrada) .....	23
<b>Slika 23:</b> Prikaz vizualizacije promjene cijene prema vrsti goriva 1.primjer. (izvor: vlastita izrada) .....	24
<b>Slika 24:</b> Prikaz vizualizacije promjene cijene prema vrsti goriva 2.primjer. (izvor: vlastita izrada) .....	25
<b>Slika 25:</b> Prikaz postupka izrade vremenskih serija cijena. (izvor: vlastita izrada) ....	26
<b>Slika 26:</b> Prikaz vizualizacije vremenske serije cijena. (izvor: vlastita izrada) .....	27
<b>Slika 27:</b> Prikaz postupka izrade vizualizacije korelacijske matrice 1.dio. (izvor: vlastita izrada) .....	28
<b>Slika 28:</b> Prikaz postupka izrade vizualizacije korelacijske matrice 2.dio. (izvor: vlastita izrada) .....	29
<b>Slika 29:</b> Prikaz vizualizacije korelacijske matrice. (izvor: vlastita izrada) .....	29
<b>Slika 30:</b> Prikaz naredbe za instalaciju TabPy dodatka (izvor: vlastita izrada) .....	31
<b>Slika 31:</b> Prikaz povezivanja alata Tableau na TabPy server (izvor: vlastita izrada) .....	32
<b>Slika 32:</b> Prikaz postupka izrade vizualizacije za razliku cijene i poreza goriva (izvor: vlastita izrada) .....	32
<b>Slika 33:</b> Prikaz vizualizacije razlike cijene goriva i poreza. (izvor: vlastita izrada) ....	33

## Prilozi

U prilogu se nalazi Python kod u datoteci sa **.ipynb** ekstenzijom koja je izvezena iz Jupyter Notebook okruženja.

# Analiza cijena goriva u Hrvatskoj, Sloveniji i Mađarskoj za 2022. godinu

## Učitavanje i čitanje podataka

```
In [36]: import pandas as pd
```

```
In [37]: df = pd.read_excel('E:/cijene/Cijene.xlsx')
```

```
In [38]: df.head()
```

```
Out[38]:
```

	Prices in force on	Country Name	Country EU Code	Product Name	Currency Code	Prices Unit	Euro exchange rate	Weekly price with taxes	Weekly price without taxes
0	2022-01-03	Croatia	HR	Automotive gas oil	HRK	1000L	0.132996	1,463.49	763.82
1	2022-01-03	Croatia	HR	Euro-super 95	HRK	1000L	0.132996	1,484.9	674.56
2	2022-01-03	Croatia	HR	Fuel oil - Sulphur less than 1%	HRK	t	0.132996	712.73	691.45
3	2022-01-03	Croatia	HR	Heating gas oil	HRK	1000L	0.132996	756.75	559.78
4	2022-01-03	Croatia	HR	LPG - motor fuel	HRK	1000L	0.132996	786.27	621.65

## Informacije o podacima

```
In [39]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 714 entries, 0 to 713
Data columns (total 9 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   Prices in force on                    714 non-null    datetime64[ns]
 1   Country Name                          714 non-null    object
 2   Country EU Code                       714 non-null    object
 3   Product Name                          714 non-null    object
 4   Currency Code                         714 non-null    object
 5   Prices Unit                           714 non-null    object
 6   Euro exchange rate                    714 non-null    float64
 7   Weekly price with taxes               714 non-null    object
 8   Weekly price without taxes            714 non-null    object
dtypes: datetime64[ns](1), float64(1), object(7)
memory usage: 50.3+ KB
```

## Transformacija i čišćenje podataka

```
In [40]: #Najprije je potrebno prebaciti stupac sa cijenama u float te ukloniti zareze u odvajanju
df['Weekly price with taxes'] = df['Weekly price with taxes'].str.replace(',', '').astype(float)
df['Weekly price without taxes'] = df['Weekly price without taxes'].str.replace(',', '').astype(float)
```

```
In [41]: #Zamjena mjerne jedinica tako da svuda piše 1000L
df['Prices Unit'] = df['Prices Unit'].replace({'t': '1000L'})
df.to_excel("E:/Podaci za tableau/Cijenetransofrmirano.xlsx", index=False)
```

```
In [42]: df.head()
```

Out[42]:

	Prices in force on	Country Name	Country EU Code	Product Name	Currency Code	Prices Unit	Euro exchange rate	Weekly price with taxes	Weekly price without taxes
0	2022-01-03	Croatia	HR	Automotive gas oil	HRK	1000L	0.132996	1463.49	763.82
1	2022-01-03	Croatia	HR	Euro-super 95	HRK	1000L	0.132996	1484.90	674.56
2	2022-01-03	Croatia	HR	Fuel oil - Sulphur less than 1%	HRK	1000L	0.132996	712.73	691.45
3	2022-01-03	Croatia	HR	Heating gas oil	HRK	1000L	0.132996	756.75	559.78
4	2022-01-03	Croatia	HR	LPG - motor fuel	HRK	1000L	0.132996	786.27	621.65

## Izračun prosječne, maksimalne i minimalne cijene

```
In [43]: # Grupiranje dataframe-a prema državi, vrsti goriva, te izračun prosječne, minimalne i m
# Funkcija za pridruživanje datuma max i min cijenama
def get_max_min_dates(group):
    max_price_idx = group['Weekly price with taxes'].idxmax()
    min_price_idx = group['Weekly price with taxes'].idxmin()
    group['Max na datum'] = group.loc[max_price_idx, 'Prices in force on']
    group['Min na datum'] = group.loc[min_price_idx, 'Prices in force on']
    return group

# Grupiranje i primjena funkcije
result = df.groupby(['Country Name', 'Product Name'], group_keys=False).apply(get_max_min

# Izračun prosječne, max i min cijene
result = result.groupby(['Country Name', 'Product Name']).agg({
    'Weekly price with taxes': ['mean', 'max', 'min'],
    'Max na datum': 'first', # Prva vrijednost nakon grupiranja je ona koja je pridruže
    'Min na datum': 'first' # Prva vrijednost nakon grupiranja je ona koja je pridruže
}).reset_index()

result.columns = ['Country Name', 'Product Name', 'Prosječna cijena', 'Max Cijena', 'Min
result.to_excel("E:/Podaci za tableau/Prosje_max_min_sve_drz.xlsx", index=False)

result
```

Out[43]:

	Country Name	Product Name	Prosječna cijena	Max Cijena	Min Cijena	Max na datum	Min na datum
0	Croatia	Automotive gas oil	1762.536078	1947.25	1463.49	2022-07-04	2022-01-03
1	Croatia	Euro-super 95	1604.532549	1873.26	1331.17	2022-06-06	2022-12-26
2	Croatia	Fuel oil - Sulphur less than 1%	922.257647	1110.35	711.89	2022-03-21	2022-01-10
3	Croatia	Heating gas oil	1188.309200	1471.53	756.75	2022-07-11	2022-01-03
4	Croatia	LPG - motor fuel	944.967255	1050.85	786.27	2022-06-13	2022-01-03

5	Hungary	Automotive gas oil	1478.166078	1778.38	1220.32	2022-12-26	2022-03-07
6	Hungary	Euro-super 95	1289.545882	1585.74	1202.99	2022-12-19	2022-10-10
7	Hungary	Fuel oil - Sulphur less than 1%	815.240980	960.96	559.97	2022-06-06	2022-01-24
8	Hungary	Heating gas oil	1478.166078	1778.38	1220.32	2022-12-26	2022-03-07
9	Hungary	LPG - motor fuel	954.940000	1022.27	839.89	2022-08-08	2022-03-07
10	Slovenia	Automotive gas oil	1635.272353	1865.51	1367.28	2022-07-04	2022-01-03
11	Slovenia	Euro-super 95	1493.366471	1772.32	1318.03	2022-07-04	2022-01-03
12	Slovenia	Heating gas oil	1326.694706	1672.12	974.76	2022-06-27	2022-01-03
13	Slovenia	LPG - motor fuel	975.042549	1054.42	833.37	2022-04-25	2022-02-07

```
In [44]: # Izračun minimalne i maksimalne cijene prema vrsti goriva
max_prices = df.groupby('Product Name')['Weekly price with taxes'].idxmax()
min_prices = df.groupby('Product Name')['Weekly price with taxes'].idxmin()

# Izračun datuma
max_price_df = df.loc[max_prices, ['Product Name', 'Country Name', 'Weekly price with ta
min_price_df = df.loc[min_prices, ['Product Name', 'Country Name', 'Weekly price with ta

max_price_df.columns = ['Product Name', 'Država sa Max Cijenom', 'Max Cijena', 'Datum Ma
min_price_df.columns = ['Product Name', 'Država sa Min Cijenom', 'Min Cijena', 'Datum Mi

#Resetiranje indexa
max_price_df.reset_index(drop=True, inplace=True)
min_price_df.reset_index(drop=True, inplace=True)

result = pd.merge(max_price_df, min_price_df, on='Product Name')

result.to_excel("E:/Podaci za tableau/Prosje_max_min_za_goriva.xlsx", index=False)
result
```

```
Out[44]:
```

	Product Name	Država sa Max Cijenom	Max Cijena	Datum Max Cijene	Država sa Min Cijenom	Min Cijena	Datum Min Cijene
0	Automotive gas oil	Croatia	1947.25	2022-07-04	Hungary	1220.32	2022-03-07
1	Euro-super 95	Croatia	1873.26	2022-06-06	Hungary	1202.99	2022-10-10
2	Fuel oil - Sulphur less than 1%	Croatia	1110.35	2022-03-21	Hungary	559.97	2022-01-24
3	Heating gas oil	Hungary	1778.38	2022-12-26	Croatia	756.75	2022-01-03
4	LPG - motor fuel	Slovenia	1054.42	2022-04-25	Croatia	786.27	2022-01-03

## Promjena cijena kroz godinu za svaku državu prema vrsti

```
In [49]: # Sortiranje podataka prema državi, vrsti goriva i datumu
df_promjena = df.sort_values(['Country Name', 'Product Name', 'Prices in force on'])

# Izračunavanje promjena cijena kroz godinu za svaku državu i vrstu goriva
df_promjena['Promjena cijene'] = df_promjena.groupby(['Country Name', 'Product Name'])['
```

In [50]: df\_promjena.head()

Out[50]:

	Prices in force on	Country Name	Country EU Code	Product Name	Currency Code	Prices Unit	Euro exchange rate	Weekly price with taxes	Weekly price without taxes	Promjena cijene
0	2022-01-03	Croatia	HR	Automotive gas oil	HRK	1000L	0.132996	1463.49	763.82	NaN
5	2022-01-10	Croatia	HR	Automotive gas oil	HRK	1000L	0.132841	1465.50	765.91	2.01
10	2022-01-17	Croatia	HR	Automotive gas oil	HRK	1000L	0.132846	1492.53	787.51	27.03
15	2022-01-24	Croatia	HR	Automotive gas oil	HRK	1000L	0.132820	1519.99	809.56	27.46
20	2022-01-31	Croatia	HR	Automotive gas oil	HRK	1000L	0.132814	1520.19	809.74	0.20

## Izračun korelacijske matrice

In [59]:

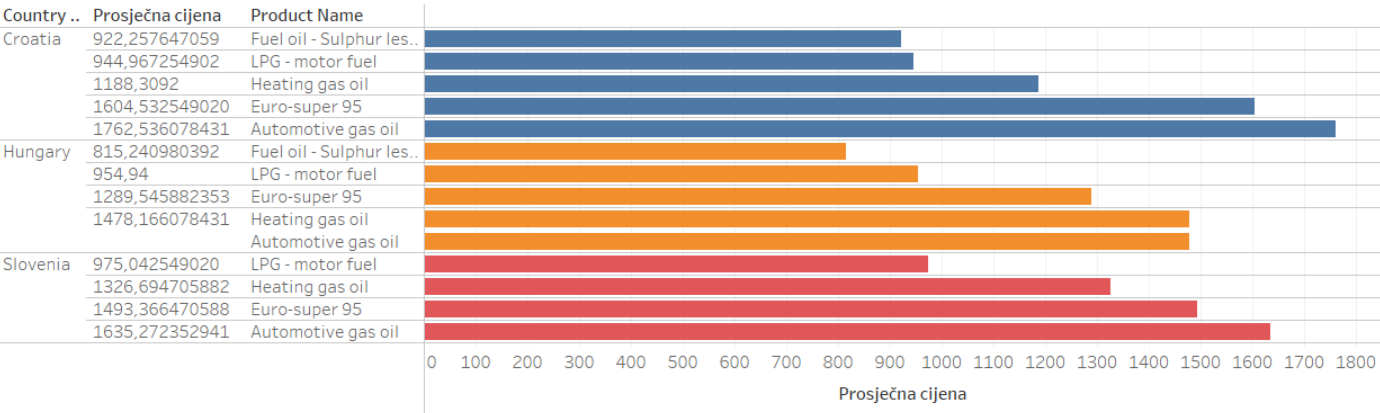
```
grupiranje_po_gorivu = df.groupby(['Product Name', 'Prices in force on'])['Weekly price']  
  
# Pivotiranje podataka tako da svaka vrsta goriva postane zaseban stupac  
pivot_df = grupiranje_po_gorivu.pivot(index='Prices in force on', columns='Product Name')  
  
# Izračun korelacijske matrice  
correlation_matrix = pivot_df.corr()  
correlation_matrix.to_excel("E:/Podaci za tableau/korelacijska_matrica.xlsx")  
correlation_matrix
```

Out[59]:

	Product Name	Automotive gas oil	Euro-super 95	Fuel oil - Sulphur less than 1%	Heating gas oil	LPG - motor fuel
	Product Name					
	Automotive gas oil	1.000000	0.351173	0.786315	0.965548	0.821554
	Euro-super 95	0.351173	1.000000	0.483631	0.441623	0.577598
	Fuel oil - Sulphur less than 1%	0.786315	0.483631	1.000000	0.803224	0.844689
	Heating gas oil	0.965548	0.441623	0.803224	1.000000	0.873595
	LPG - motor fuel	0.821554	0.577598	0.844689	0.873595	1.000000

In [ ]:

Prosječna cijena goriva za sve države za svaku vrstu goriva



Sum of Prosječna cijena for each Product Name broken down by Country Name and Prosječna cijena. Color shows details about Country Name.

- Country Name
- Croatia
  - Hungary
  - Slovenia



## Goriva s maksimalnom cijenom kroz godinu

Država sa Max Cijenom	Product Name	Day of Datum Max Cijene				
		21. ožujka 2022.	25. travnja 2022.	6. lipnja 2022.	4. srpnja 2022.	26. prosinca 2022.
Croatia	Automotive gas oil				1.947,3	
	Euro-super 95			1.873,3		
	Fuel oil - Sulphur less than 1%	1.110,4				
Hungary	Heating gas oil					1.778,4
Slovenia	LPG - motor fuel		1.054,4			

Sum of Max Cijena broken down by Datum Max Cijene Day vs. Država sa Max Cijenom and Product Name. Color shows details about Država sa Max Cijenom. The marks are labeled by sum of Max Cijena.

- Država sa Max Cijenom
- Croatia
  - Hungary
  - Slovenia

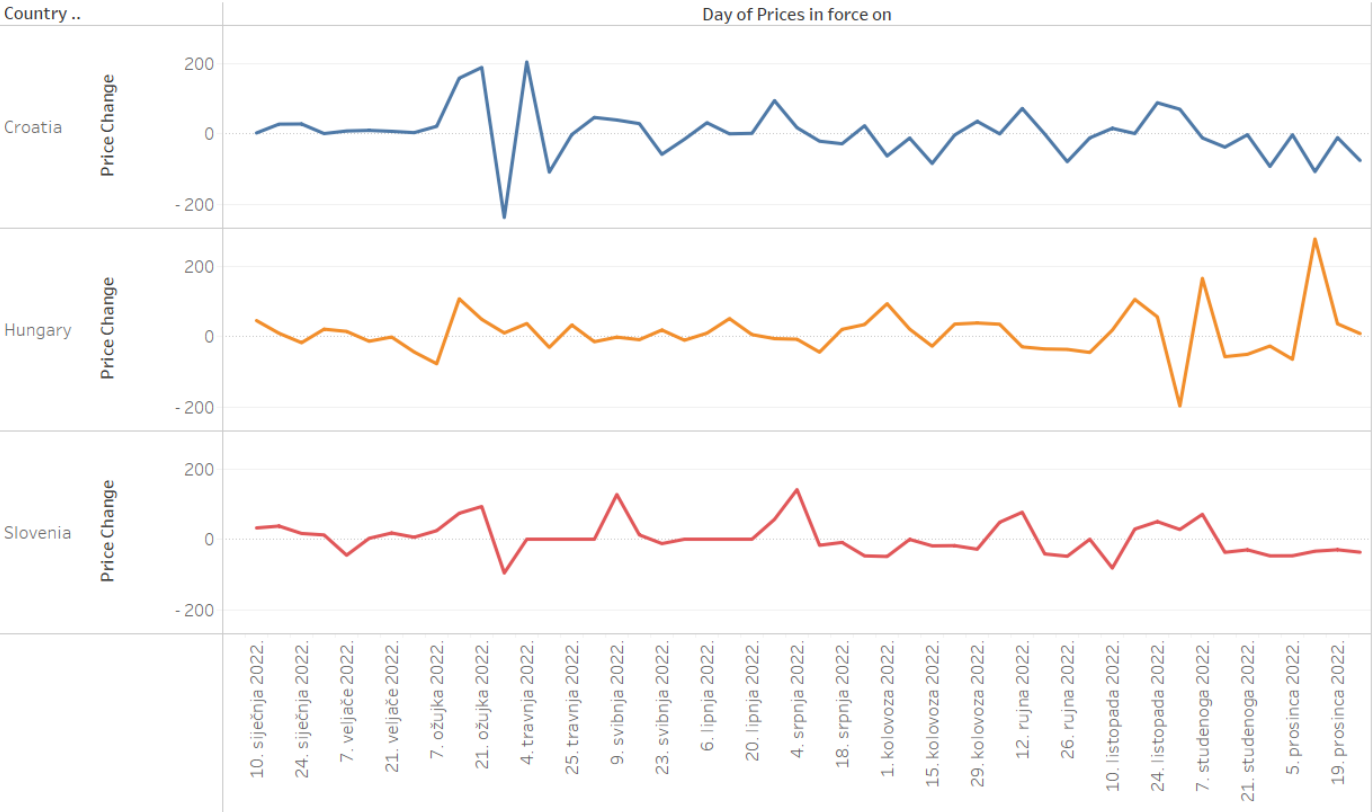
## Goriva s minimalnom cijenom kroz godinu

Država sa Min Cijenom	Product Name	Day of Datum Min Cijene			
		3. siječnja 2022.	24. siječnja 2022.	7. ožujka 2022.	10. listopada 2022.
Croatia	Heating gas oil	756,8			
	LPG - motor fuel	786,3			
Hungary	Automotive gas oil			1.220,3	
	Euro-super 95				1.203,0
	Fuel oil - Sulphur less than 1%		560,0		

Sum of Min Cijena broken down by Datum Min Cijene Day vs. Država sa Min Cijenom and Product Name. Color shows details about Država sa Min Cijenom. The marks are labeled by sum of Min Cijena.

- Država sa Min Cijenom
- Croatia
  - Hungary

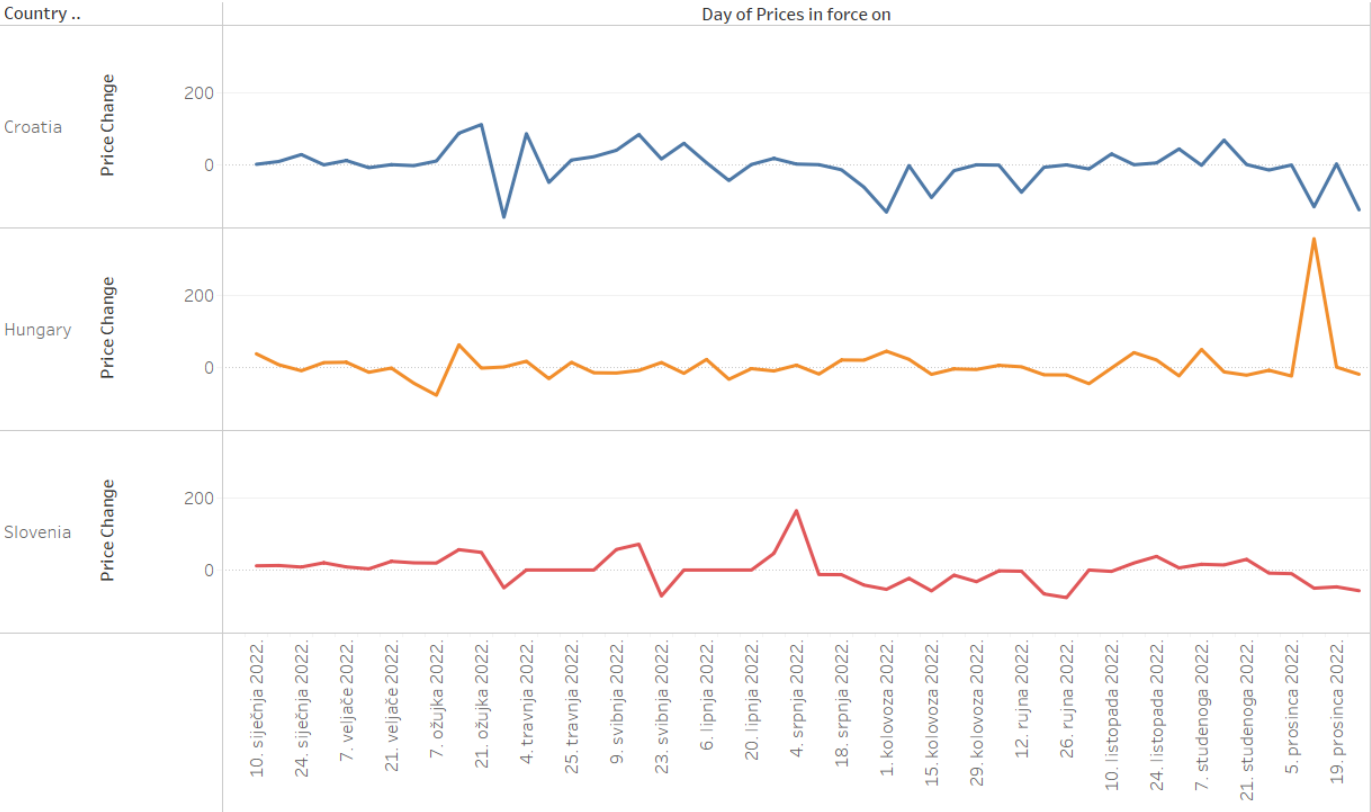
Promjena cijena po državi za Automotive gas oil



The trend of sum of Price Change for Prices in force on Day broken down by Country Name. Color shows details about Country Name. The data is filtered on Product Name, which keeps Automotive gas oil.

- Country Name
- Croatia
  - Hungary
  - Slovenia

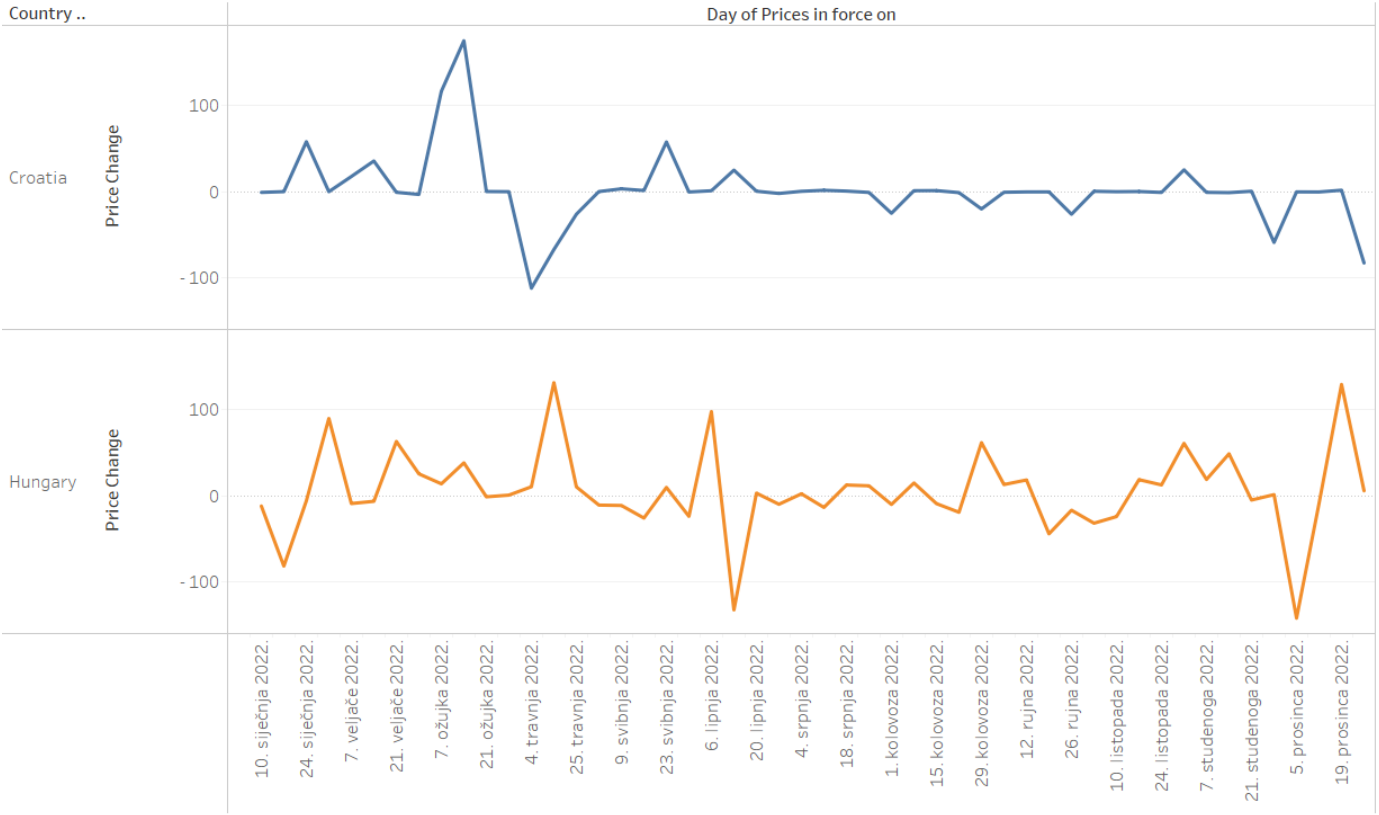
Promjena cijena po državi za Euro-super 95



The trend of sum of Price Change for Prices in force on Day broken down by Country Name. Color shows details about Country Name. The data is filtered on Product Name, which keeps Euro-super 95.

- Country Name
- Croatia
  - Hungary
  - Slovenia

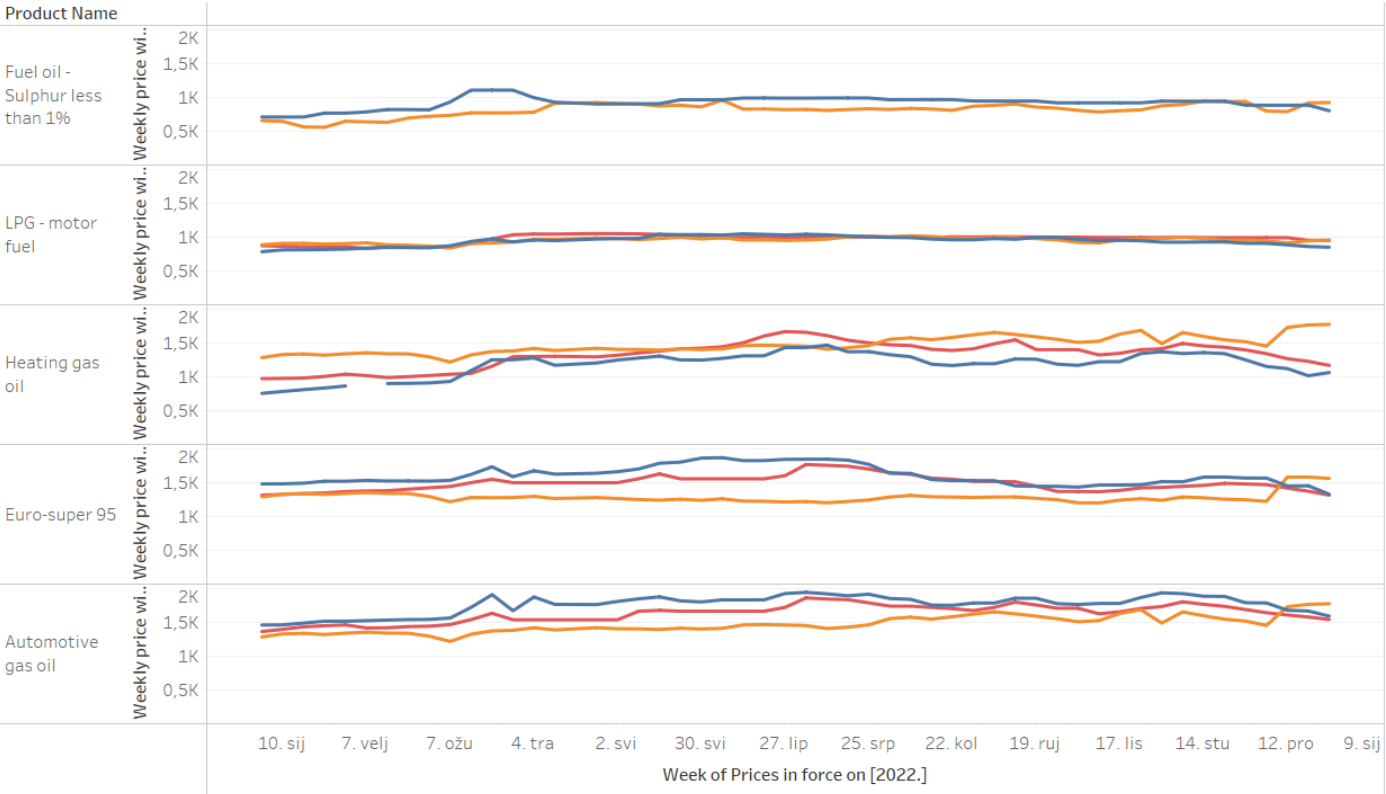
Promjena cijena po državi za Fuel oil - Sulphur less than 1%



The trend of sum of Price Change for Prices in force on Day broken down by Country Name. Color shows details about Country Name. The data is filtered on Product Name, which keeps Fuel oil - Sulphur less than 1%.

- Country Name
- Croatia
  - Hungary

Vremenska serija cijena



The trend of sum of Weekly price with taxes for Prices in force on Week broken down by Product Name. Color shows details about Country Name.

Country Name

- Croatia
- Hungary
- Slovenia

Korelacijska Matrica

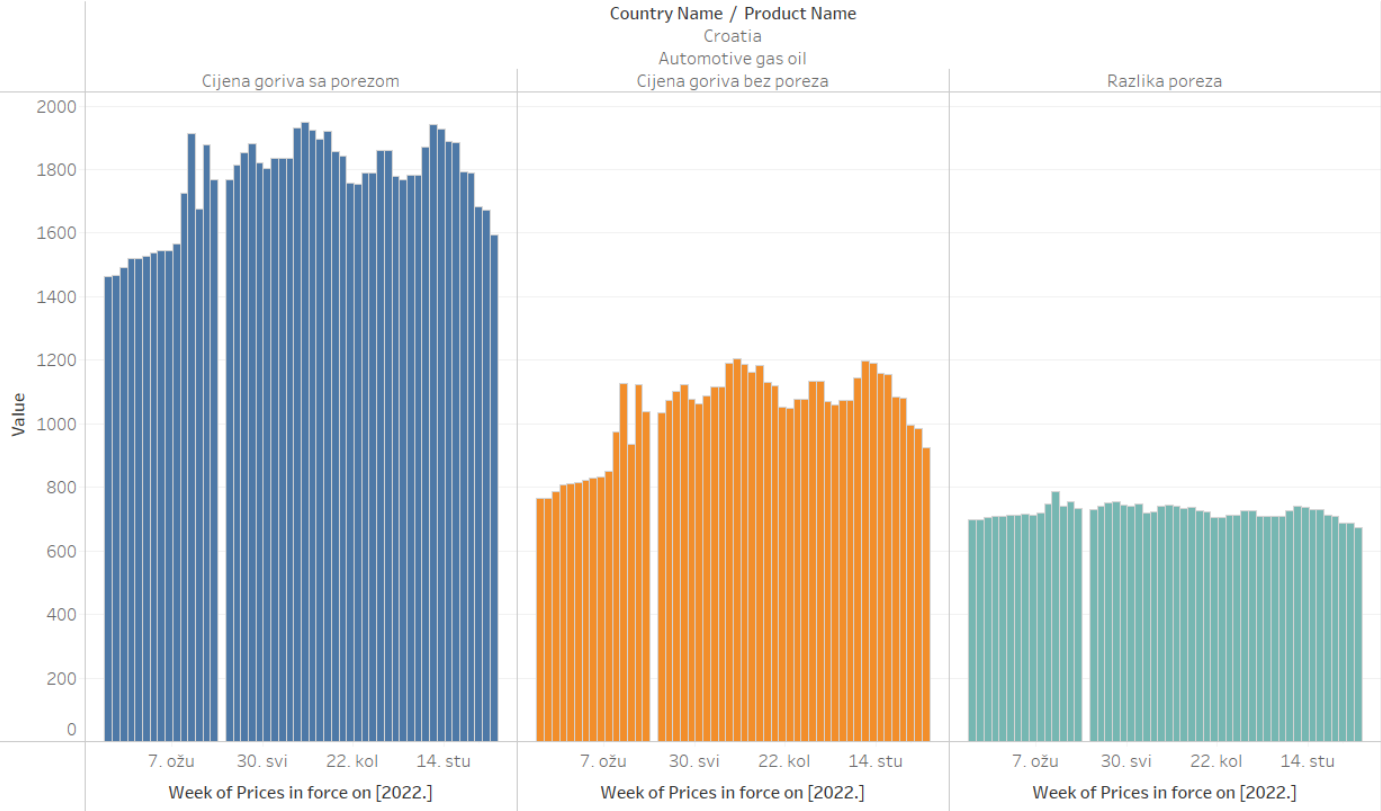
Product Name	Automotive gas oil	Euro-super 95	Fuel oil - Sulphur les..	Heating gas oil	LPG - motor fuel
Automotive gas oil	1,0000	0,3512	0,7863	0,9655	0,8216
Euro-super 95	0,3512	1,0000	0,4836	0,4416	0,5776
Fuel oil - Sulphur les..	0,7863	0,4836	1,0000	0,8032	0,8447
Heating gas oil	0,9655	0,4416	0,8032	1,0000	0,8736
LPG - motor fuel	0,8216	0,5776	0,8447	0,8736	1,0000

Automotive gas oil, Euro-super 95, Fuel oil - Sulphur less than 1%, Heating gas oil and LPG - motor fuel broken down by Product Name. Color shows Automotive gas oil, Euro-super 95, Fuel oil - Sulphur less than 1%, Heating gas oil and LPG - motor fuel. The marks are labeled by Automotive gas oil, Euro-super 95, Fuel oil - Sulphur less than 1%, Heating gas oil and LPG - motor fuel.

Measure Values

0,3512 1,0000

Razlika cijena goriva bez poreza za Automotive Gas Oil u Hrvatskoj



The plots of Cijena goriva sa porezom, Cijena goriva bez poreza and Razlika poreza for Prices in force on Week broken down by Country Name and Product Name. Color shows details about Cijena goriva sa porezom, Cijena goriva bez poreza and Razlika poreza . The view is filtered on Country Name and Product Name. The Country Name filter keeps Croatia. The Product Name filter keeps Automotive gas oil.

- Measure Names
- Cijena goriva sa porezom
  - Cijena goriva bez poreza
  - Razlika poreza