



Univerzitet u Nišu  
Elektronski fakultet



**Razvoj deskriptivnih modela odlučivanja**  
**Segmentacija klijenata i analiza potrošačke korpe**  
Poslovna inteligencija  
Seminarski rad

Student:

Barbara Mitić, br.indeksa 1639

Mentor:

Doc. dr Aleksandar Stanimirović

Niš, 2024.

# Sadržaj

1. Uvod.....	1
2. Teorijska osnova .....	2
2.1 Poslovna inteligencija .....	2
2.1.1 Arhitektura poslovne inteligencije .....	3
2.2 Poslovna analitika .....	5
2.3 Segmentacija klijenata .....	6
2.3.1 Tipovi segmentacije kupaca .....	7
2.3.2 RFM model .....	11
2.3.3 CLV model .....	16
2.3.4 Metod klasterizacije .....	19
2.4 Analiza potrošačke korpe .....	22
2.4.1 Algoritmi za analizu potrošačke korpe .....	24
3. Detalji implementacije .....	27
4. Zaključak .....	42
5. Literatura .....	43

# 1. Uvod

U današnje vreme poslovno okruženje postaje sve složenije i dinamičnije, a kompanije i ostala preduzeća, bez obzira na veličinu i sektor, suočavaju se sa sve većim izazovima da ostanu konkurentni na tržištu i uspešni. Svaka kompanija teži da se prilagodi promenama i donese ispravne odluke za dalje poslovanje. To zahteva pažljivo odabranu strategiju za brzo planiranje i donošenje odluka, a samim tim i precizne i pravovremene informacije. Razvoj digitalnih tehnologija, široka upotreba interneta i dostupnost informacija, doveli su do lakog pristupa velikoj količini podataka. Međutim, ovako raznovrsni i obimni podaci često su nesređeni i teško razumljivi, što otežava njihovu direktnu primenu u poslovnom odlučivanju. Da bi prikupljeni podaci bili korisni, potrebno ih je pravilno obraditi i analizirati. U tu svrhu je razvijena poslovna inteligencija.

Poslovna inteligencija (eng. *Business Intelligence, BI*) se može definisati kao skup alata i tehnologija za prikupljanje, obradu i analizu podataka i njihovo “pretvaranje” u korisne informacije koje pomažu kompanijama u donošenju poslovnih odluka. Cilj poslovne inteligencije je da omogući kompanijama da bolje razumeju svoje poslovanje, identifikuju prilike i rizike, kao i da povećaju efikasnost i produktivnost. Kroz analizu podataka *BI* pomaže u otkrivanju obrazaca u ponašanju klijenata i trendova, podstiče efikasnije upravljanje resursima i donošenje ispravnih strateških odluka zasnovanih na činjenicama, a ne na pretpostavkama [1].

Dok *BI* obuhvata širu sliku – uključujući prikupljanje podataka, njihovo skladištenje, obradu i prezentaciju, poslovna analitika se usmerava na analitičke metode i alate koji se koriste za detaljnije ispitivanje tih podataka. Analiza u okviru poslovne inteligencije obuhvata različite pristupe, uključujući deskriptivnu, dijagnostičku, prediktivnu i preskriptivnu analizu. Deskriptivna analiza pruža uvid u ono što se desilo u kompaniji i pomaže identifikovanju ključnih trendova i obrazaca ponašanja. Dijagnostička analiza identifikuje faktore i uzroke koji su doveli do određenih poslovnih rezultata. Prediktivna analiza koristi podatke i statističke modele za predviđanje budućih događaja ili trendova, dok preskriptivna analiza ide korak dalje, preporučujući konkretne akcije koje treba preduzeti za postizanje željenih poslovnih rezultata.

U fokusu ovog rada je deskriptivna analiza, koja igra ključnu ulogu u razumevanju ponašanja potrošača i donošenju poslovnih odluka zasnovanih na podacima. Posebna pažnja posvećena je segmentaciji klijenata i analizi potrošačke korpe. Detalji implementacije i analiza dobijenih rezultata prikazani su u poglavlju 3. Zaključci izvedeni iz istraživanja nalaze se u poglavlju 4, dok je u poglavlju 5 dat spisak korišćene literature.

## 2. Teorijska osnova

Termini poslovna inteligencija (eng. *Business Intelligence*) i poslovna analitika (eng. *Business Analytics*) se često koriste kako sinonimi, što dovodi do stalnih rasprava o tome da li je poslovna analitika podskup poslovne inteligencije ili obratno. Iako se ove dve oblasti preklapaju u mnogim aspektima, one se bitno razlikuju po svom fokusu i pristupu analizi podataka.

### 2.1 Poslovna inteligencija

Izraz *poslovna inteligencija* je prvi upotrebio IBM-ov istraživač Hans Peter Lun. U članku napisanom 1958. definisao ju je kao “sposobnost zapažanja veza među prezentovanim podacima na način koji pomaže aktivnostima koje vode ka određenom cilju”.

*Poslovna inteligencija* je skup metodologija i softverskih alata za identifikaciju, prikupljanje i analizu poslovnih podataka koji omogućavaju njihovo efikasno korišćenje, najčešće iz skladišta podataka i pretvaranje u informacije potrebne za donošenje poslovnih odluka. Cilj poslovne inteligencije je da se iz velikog skupa internih i eksternih podataka preduzeća, uočavanjem njihovih međusobnih veza i zakonitosti pojavljivanja, dobiju upravljačke informacije kojima se povećava uspešnost poslovanja [2].

Za uspešno poslovanje potrebno je donositi ispravne odluke, što zahteva upotrebu *BI* sistema zasnovanih na savremenim tehnologijama. Bez poslovne inteligencije kompanije ne mogu u potpunosti iskoristiti prednosti koje pruža donošenje odluka zasnovanih na analizi podataka. U tom slučaju se važne poslovne odluke zasnivaju na faktorima kao što su: akumulirano znanje, prethodno iskustvo, intuicija i osećanja.

Vremenom se poslovna inteligencija sve više razvijala i sada uključuje više procesa i aktivnosti, među kojima su [5]:

- Istraživanje podataka (eng. *data mining*) – korišćenje baza podataka, statistike i mašinskog učenja za otkrivanje trendova u velikim skupovima podataka
- Izveštavanje (eng. *reporting*) – deljenje analiza podataka sa interesnim stranama radi izvođenja zaključaka i donošenja odluka
- Merenje performansi (eng. *performance metrics*) – upoređivanje trenutnih podataka o učinku sa prethodnim podacima za praćenje napretka u postizanju zadatih ciljeva, obično korišćenjem kontrolnih tabli (eng. *dashboard*)
- Postavljanje upita (eng. *querying*) – postavljanje pitanja specifičnih za podatke i izvlačenje odgovora iz skupova podataka
- Statistička analiza – korišćenje rezultata deskriptivne analize za dalju statističku obradu
- Vizualizacija podataka - vizuelni prikaz rezultata analize podataka

### 2.1.1 Arhitektura poslovne inteligencije

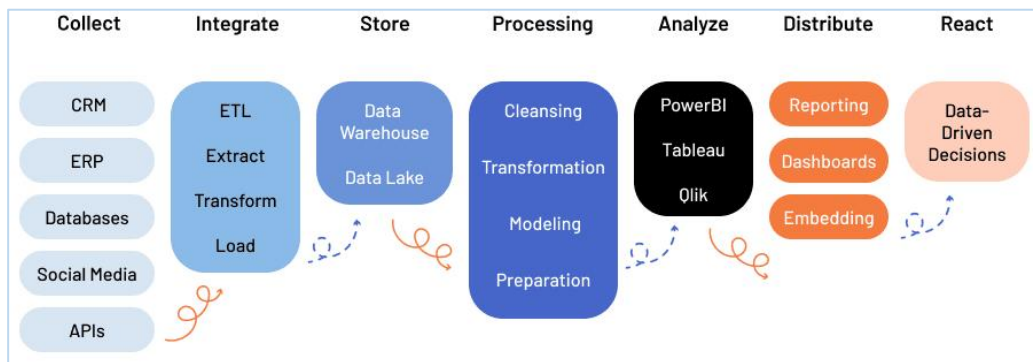
*BI* arhitektura određuje tehnološke standarde, standarde za upravljanje podacima i analitiku koji podržavaju poslovne inicijative u jednoj kompaniji, kao i konkretne platforme i alate koji će u tu svrhu biti korišćeni. *BI* arhitektura zapravo definiše plan kako će kompanije sprovesti svoje poslovne aktivnosti. Bez adekvatne *BI* arhitekture, kompanije rizikuju da naprave skupe greške dok pokušavaju da iskoriste raspoložive podatke. Međutim, uz dobro osmišljenu *BI* arhitekturu, kompanija može lako izbeći te prepreke i postaviti temelje za uspeh i napredak.

Arhitekturu *BI* sistema čine tri komponente [3]:

- Upravljanje podacima (eng. *Data management*) – odnosi se na način prikupljanja podataka, njihovo skladištenje i pristupanje. Upravljanje podacima postoji na svim nivoima u kompaniji, počev od podataka prikupljenih preko senzora i ostalih uređaja, sve do podataka koje zaposleni unose u baze podataka. Takođe, podaci se prikupljaju i iz eksternih izvora poput društvenih mreža, marketinških istraživanja i sl. Zadatak ove komponente je da obezbedi tačnost, ažurnost i jednostavno pristupanje podacima.
- Analitika (eng. *Analytics*) – odnosi se na pretvaranje sirovih podataka u korisne informacije tj. uvide. Analiza podataka pomaže kompanijama da bolje razumeju svoje klijente, poslovanje i trendove u industriji. U praksi postoje različite vrste analitika, ali sve imaju isti cilj – pomoći kompanijama u donošenju dobrih poslovnih odluka.
- Tehnologija (eng. *Technology*) – odnosi se na hardver i softver koje kompanije koriste kako bi podržale svoje *BI* aktivnosti. To uključuje sve, počev od skladišta podataka, preko ETL alata, do platformi za izveštavanje i vizuelizaciju.
  - Skladišta podataka (eng. *Data warehouses*) – baza podataka dizajnirana za skladištenje i analizu podataka. Kompanije koriste skladišta podataka kako bi svoje podatke čuvale na jednom mestu, gde im se lako može pristupiti i analizirati.
  - Jezera podataka (eng. *Data lakes*) – repozitorijum podataka koji se takođe može koristiti za čuvanje podataka i njihovu analizu. Kompanije ih obično koriste za skladištenje nestruktuiranih podataka koji još uvek nisu spremni za analizu.
  - Alati poslovne inteligencije (eng. *Business Intelligence Tools*) – softverske aplikacije za prikupljanje, skladištenje i analizu podataka. Ovi alati se takođe koriste za kreiranje izveštaja, kontrolnih tabli i ostalih vizuelizacija koje pomažu kompanijama da bolje razumeju svoje podatke.

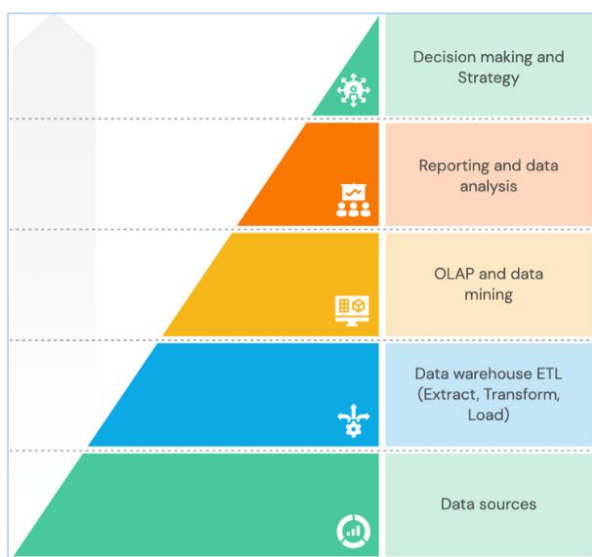
Kao što je već pomenuto, izvori podataka obuhvataju interne izvore (baze podataka, ERP sistemi, CRM sistemi) i eksterne izvore (društvene mreže, marketinška istraživanja, web analitika). Da bi se prikupljeni podaci pravilno integrisali u *BI* ekosistem, koriste se alati za ekstrakciju, transformaciju i učitavanje (eng. *Extract/Transform/Load, ETL*). ETL alati su: Azure Data Factory, Google Cloud Dataflow, AWS Glue, Talend itd. Podaci se dalje smeštaju u skladište ili jezero podataka. Podacima se pristupa radi obrade: čišćenje, transformacija, modelovanje i priprema za analizu. Obradeni podaci se analiziraju korišćenjem alata kao što PowerBI, Tableau ili Qlik. Rezultati analize se distribuiraju kroz izveštaje, kontrolne table ili

vizuelizacije. Na osnovu rezultata analize donose se odluke zasnovane na podacima kako bi se unapredile poslovne operacije. Na slici 1 ilustrovana je tipična arhitektura *BI* sistema.



**Slika 1.** Arhitektura *BI* sistema

U literaturi se struktura *BI* sistema često prikazuje u obliku piramide (slika 2.) Prikupljeni podaci se, nakon skladištenja, istražuju i analiziraju. Alati koji se na ovom nivou koriste formiraju tzv. pasivnu *BI* analizu, sačinjenu od upita, sistema izveštavanja i statističkih modela. Ove metode se nazivaju pasivnim zato što od donosioca odluke zahtevaju da prethodno postave hipoteze ili definišu kriterijume za ekstrakciju podataka, a tek onda se koriste alati za analizu. Na sledećem nivou se vrši otkrivanje znanja, tj. izdvajanje korisnih informacija iz sirovih podataka. U tu svrhu se koriste matematički modeli za prepoznavanje obrazaca, mašinsko učenje i *data mining* tehnike. Modeli na ovom nivou se nazivaju aktivnim zato što ne zahtevaju od donosioca odluke da prethodno postavi hipotezu koju je potrebno proveriti. Naredni nivo u piramidi je rezervisan za optimizacione modele – u ovom koraku se biraju najbolje rešenja ili alternativne strategije na osnovu rezultata prethodnih analiza i znanja otkrivenog iz podataka. Na vrhu piramide se donose poslovne odluke na osnovu svih prethodnih nivoa analize. Kompanije, tj donosioci odluka korišćenjem neformalnih i nestruktuiranih podataka mogu da utiču na dobijene rezultate tj. da izmene dobijene preporuke [3].



**Slika 2.** Komponente *BI* sistema

## 2.2 Poslovna analitika

Poslovna analitika je proces koji uključuje korišćenje podataka, statističkih modela i tehnologija za analizu poslovnih informacija i donošenje informisanih poslovnih odluka. Cilj poslovne analitike je prepoznavanje obrazaca, uvida i otkrivanje trendova koji pomažu kompanijama da optimizuju poslovne procese, povećaju efikasnost, smanje troškove i unaprede konkurentsku prednost [6].

**Deskriptivna analitika** – ovaj nivo analitike se fokusira na analizu već prikupljenih podataka kako bi se stekao uvid u to šta se desilo u prošlosti. Deskriptivna analitika se ne koristi za izvođenje zaključaka ili predviđanje. Kao najjednostavniji oblik analize podataka, koristi jednostavne matematičke i statističke metode, poput aritmetičke sredine i procentualnih promena, umesto složenih proračuna potrebnih za naprednije tipove analitike. Za prikazivanje podataka se koriste jednostavne vizuelizacije poput linijskih (eng. *line chart*) i pita (eng. *pie chart*) grafikona, kao i kontrolne table (eng. *dashboard*) koje pružaju objedinjeni vizuelni prikaz ključnih informacija. Uobičajeni primeri deskriptivne analitike su izveštaji koji pružaju “istorijske” uvide u vezi sa proizvodnjom, finansijama, operacijama, inventarima i klijentima neke kompanije [7].

**Dijagnostička analitika** – istražuje podatke sa ciljem razumevanja uzroka određenih postupaka, događaja i rezultata. Analitičari koriste različite tehnike i alate za identifikaciju obrazaca, trendova i veza koje objašnjavaju zašto su se određeni događaji desili. Glavni cilj dijagnostičke analitike je da pruži uvid u faktore koji su uticali na specifične ishode. Dijagnostička analitika služi kao most između prošlosti i budućnosti, povezujući deskriptivnu analizu koja se bavi pitanjem “šta se dogodilo” s prediktivnom analizom koja istražuje “šta bi moglo da se dogodi”. Ona dodaje kontekst podacima pružajući detaljne uvide koji pomažu kompanijama da donesu preciznije odluke i bolje razumeju posledice prethodnih odluka. Dijagnostička analitika koristi tehnike poput regresione analize, analize uzročno-posledičnih veza, i analiza podataka o izuzetnim slučajevima [7].

**Prediktivna analitika** – predviđa šta će se desiti u budućnosti. Bazirana je na statističkim tehnikama, kao i na metodama *data mining*-a. Među najčešće korišćenim tehnikama su algoritmi klasifikacije kao što su logistička regresija, modeli stabala odlučivanja i neuronske mreže. Pored pomenutih tehnika, koristi se i analiza vremenskih serija, klasterizacija i analiza velikih podataka (eng. *big data*). Ovaj tip analitike se primenjuje u različite svrhe, od predviđanja ponašanja kupaca i otkrivanja obrazaca kupovine, do identifikovanja trendova prodaje. Iako se prediktivna analitika oslanja na probabilističke metode i ne može garantovati potpunu tačnost, omogućava kreiranje efektivnih poslovnih strategija zasnovanih na verovatnoći i riziku [7].

**Preskriptivna analitika** – njen osnovni cilj je da pruži preporuke o tome šta treba učiniti kako bi se postigli željeni rezultati. To može uključivati preporuke o strategijama, resursima ili operativnim planovima. Koristi napredne tehnike kao što su optimizacija, simulacija i analiza scenarija. Optimizacioni algoritmi kao što su linearno i nelinearno programiranje i simulacijski modeli pomažu u identifikaciji najboljih odluka i strategija. Uključuje i tehnike kao što su analiza odluka, planiranje resursa, modeliranje scenarija i preporučivanje akcija. Na primer,

može se koristiti za preporuku cena proizvoda, određivanje strategija snabdevanja, ili optimizaciju marketinških kampanja [7].

Tipovi poslovne analitike			
<i>Deskriptivna</i>	<i>Dijagnostička</i>	<i>Prediktivna</i>	<i>Preskriptivna</i>
Koristi istorijske podatke	Koristi istorijske podatke	Koristi istorijske podatke	Koristi istorijske podatke
Identifikuje anomalije u podacima	Prevodi podatke u lako čitljiv format	Popunjava praznine u podacima	Procenjuje ishode na osnovu promenljivih
Ističe trendove u podacima	Opisuje trenutno stanje poslovanja	Kreira modele podataka	Nudi predloge o rezultatima
Investigates underlying issues	Uči iz prošlosti	Predviđa potencijalne buduće rezultate	Koristi AI i ML algoritme
„Zašto?”	„Šta?”	„Šta se može desiti?”	„Ako, onda”

**Tabela 1.** Pregled tipova poslovne analitike

## 2.3 Segmentacija klijenata

Segmentacija kupaca (eng. *Customer segmentation*) je analitički proces kojim marketinški stručnjaci “dele” tržište tj. grupišu postojeće i potencijalne kupce na osnovu zajedničkih karakteristika. Ovaj proces doprinosi boljem razumevanju njihovih potreba, preferencija i obrazaca kupovine, što marketinškim i prodajnim timovima omogućava da preduzmu adekvatne akcije sa ciljem povećanja lojalnosti kupaca [8]. Neke od prednosti segmentacije kupaca su:

- Identifikuju se najisplativiji i najmanje isplativi kupci
- Marketing i prodajni timovi se usredsređuju na kupce/klijente koji će najverovatnije kupiti proizvode/usluge kompanije
- Izbegavaju se tržišta neprofitabilna po kompaniju
- Uspostavljaju se lojalni odnosi sa kupcima/klijentima poboljšavajući proizvode/usluge kako bi se zadovoljile njihove potrebe
- Identifikuju se potencijalno novi proizvodi/usluge predviđanjem zahteva kupaca/klijenata.

Segmentacija kupaca pomaže boljem razumevanju kupaca, prepoznavanju njihovih karakteristika i zahteva, kao i definisanju daljih marketinških i prodajnih aktivnosti. Na ovaj



način se dobija uvid u interese, potrošačke navike, budžete kupaca. Česta interakcija kompanije sa kupcima podstiče njihovu lojalnost, što ih “drži vezanim” za brend kompanije.

U praksi se često koristi i segmentacija tržišta (eng. *Market segmentation*). Segmentacija tržišta je generalniji pristup i obuhvata čitavo tržište, dok se segmentacija kupaca odnosi samo na određeni deo tržišta.

Na primer, ukoliko se bavite prodajom vozila i ta vozila obično direktno prodajete nekim preduzećima/kompanijama, tada je vaš segment kupaca poslovanje-sa-poslovanjem (eng. *Business to Business, B2B*). U ovom slučaju, možete da segmentirate svoje kupce u dve grupe: one koji kupuju velike kamione za komercijalne svrhe i male preduzetnike koji kupuju kombije. Svaka grupa ima specifične potrebe, pa je korisno prilagoditi svoj pristup kako biste im najbolje odgovorili. Tako ove grupe postaju dva različita segmenta kupaca na koje možete usmeriti svoje marketinške i prodajne napore. Ukoliko pak posmatrate šire tržište, možete razmatrati kupce koji su zainteresovani za sedane (tip automobile koji se karakteriše zatvorenom karoserijom, sa četiri vrata i odvojenim prtljažnikom) u poređenju sa onima koji žele sportske automobile. Ovo je šira perspektiva, ali većina proizvođača ne može da zadovolji potrebe celokupnog tržišta. Zato je bolje fokusirati se na nekoliko segmenata kupaca, jer će to kompaniji doneti bolji rezultat u odnosu na pokušaj da se privuče celo tržište širim, generalizovanim pristupom. Ciljanje specifičnih segmenata često dovodi do veće uspešnosti i efikasnije upotrebe resursa [10].

### 2.3.1 Tipovi segmentacije kupaca

Segmentacija kupaca se obično vrši na osnovu geografskih, demografskih, psihografskih i bihevioralnih podataka.

**Geografska segmentacija** (eng. *Geographic segmentation*) u svom najjednostavnijem obliku se odnosi na mesto u kojem klijenti žive (obično grad ili država), ali može da pruži uvid u njihovo lokalno područje, uključujući klimu, vremensku zonu, kulturne preferencije i jezik. Relativno je jednostavna za implementaciju u marketinšku strategiju budući da su podaci poput kućne adrese i lokacije lako dostupni i analiziraju se bez većih problema. Često se koristi za istraživanje tržišnog potencijala u određenoj oblasti ili optimizaciju distribucionih kanala. Takođe, ona pomaže u razumevanju regionalnih trendova i preferencija, što je od ključne važnosti za kompanije koje planiraju proširenje na nova tržišta. Važno je napomenuti da geografska segmentacija daje uvid u to *gde* se nalazi vaša publika, ali ne otkriva *zašto* se potrošači ponašaju na određeni način. Iako može biti korisno sredstvo u marketinškim strategijama, često je potrebno kombinovati je sa drugim metodama segmentacije kako bi se dobila potpuna slika. Ova vrsta segmentacije ima i svoje nedostatke. Jedan od nedostataka je taj što se ovim pristupom često zanemaruju kulturne preference i kompleksnost ponašanja potrošača; činjenica da dve osobe žive na istoj adresi ne znači da imaju iste potrebe ili interesovanja [9].

Lokacija – ukoliko kompanija raspolaže informacijama o lokacijama klijenata, ona to može iskoristiti za slanje ciljanih ponuda ili poruka klijentima. Na primer, ukoliko pretražujete restorane u Parizu, tada će rezultati Google pretrage koristiti vašu lokaciju (dobijenu preko IP adrese uređaja) za pretraživanje najbližih restorana. Takođe, klijentima sa različitih geografskih

područja se nude različiti proizvodi u skladu sa preferencama lokalnog stanovništva. Na primer, u Americi se *Nike* često pojavljuje u reklamama za američki fudbal ili bejzbol, dok u u reklamama za evropski fudbal (eng. *soccer*).

Klimatski faktori – obuhvata prodaju proizvoda prilagođenih klimi, vremenskim uslovima ili sezoni u određenom geografskom području. Na primer, ako planirate putovanje na Antarktiku, topla odeća i čizme su neophodne. Zbog toga, ako ste kompanija poput *Cool Antarctica* koja proizvodi zimsku obuću, vaša ciljna grupa biće ljudi koji traže ponude za odmor u toj regiji.

Kulturološki faktori – Mc Donald's, poznati lanac restorana brze hrane, uzima u obzir kulturološke razlike, pa tako, na primer, u Indiji ne služi govedinu niti svinjetinu ni u kakvom obliku, već kao zamenu nudi specijalne indijske vegetarijanske burgere poštujući na taj način verska uverenja koja su duboko ukorenjena u indijskoj kulturi.

Primeri upotrebe geografske segmentacije:

- Proizvođači opreme za bazene ciljaju sunčana područja sa toplijom klimom za tržišta.
- Prodavnice odeće usklađuju svoj inventar sa lokalnim vremenskim uslovima i stilovima stanovnika u tom području.
- Restorani usklađuju svoje jelovnike sa lokalnim ukusima i dostupnim sastojcima u određenim područjima.
- Kompanije koje proizvode sisteme za bezbednost domaćinstava, poput alarma, ostvaruju veći uspeh u oblastima sa višom stopom kriminala.
- Maloprodajne kompanije imaju veću šansu za uspeh u područjima sa većom gustinom naseljenosti.

**Demografska segmentacija** (eng. *Demographic segmentation*) klasifikuje kupce na osnovu zajedničkih karakteristika kao što su pol, starost, bračni status, broj dece, nivo obrazovanja, zanimanje, prihodi... Ovim tipom segmentacije se lako kreira profil idealnog kupca, poboljšava se personalizacija i relevantnost proizvoda. Na primer, proizvođač automobila može da iskoristi ove podatke za ciljanje potencijalnih kupaca automobilom koji najviše pogoduje njihovom načinu života. Ukoliko kupac ima decu, verovatno želi veći, porodični automobil. Alternativno, ukoliko kupac nema neke velike prihode, malo je verovatno da bi bio zainteresovan za kupovinu vrhunskog automobila. Većina stručnjaka smatra da je ovaj tip podataka o kupcima najjednostavnije prikupiti putem anketa. Prilikom formiranja anketa, treba imati u vidu da se neka pitanja "privatnija" i osetljivija od ostalih – ljudi možda neće želeći da podele informacija o godinama, etničkoj pripadnosti ili religiji. Shodno tome, dobro je uključiti opciju "ne želim da odgovorim" kako se ispitanici ne bi našli u iskušenju da unesu netačne informacije jer ne žele da daju pravi odgovor [9]. Neka od najčešće postavljanih pitanja su:

- Koji je najviši nivo obrazovanja koji ste završili? (osnovna škola, srednja škola, visoko obrazovanje, ne želim da odgovorim)
- Koji od sledećih izraza najbolje opisuje vaš trenutni bračni status? (samac, u braku ili veren, razveden, udovac, ne želim da odgovorim)

- Prosečni prihod na nivou domaćinstva u prethodnoj godini
  - < \$10,000
  - \$10,000 do \$19,999
  - \$100,000 do \$149,000
  - >\$150,000
  - Ne želim da odgovorim

Jedan od popularnijih alata za ovaj tip analize je [Qualtrix BrandXM™](#).

**Psihografska segmentacija** (eng. *Psychographic segmentation*) se temelji na interesovanjima, vrednostima i ličnim osobinama kupaca. Za razliku od geografskih i demografskih faktora, psihografske karakteristike je teže prepoznati zato što su subjektivne i mogu značajno varirati među pojedincima unutar iste demografske grupe. Interesovanja kupaca obuhvataju različite aktivnosti kao što su sport, čitanje, umetnost ili ljubav prema životinjama, dok se vrednosti kupaca otkrivaju kroz ankete ili intervju, što olakšava kompanijama da prilagode svoje proizvode i usluge kako bi ispunile specifične potrebe kupaca. Kombinovanjem demografskih i psihografskih informacija se dobija kompletni, realni profil kupca [9].

Primer. Demografski i psihografski profil Meredith, ličnog trenera koja živi i radi u Americi.

- Demografski profil
  - Pol: Ženski
  - Starost: 25-45
  - Bračni status: U braku
  - Broj dece: 2
  - Godišnji prihod: \$75k+
  - Mesto stanovanja: Grad
- Psihografski profil
  - Voli zdrav način života
  - Voli da izlazi sa prijateljima
  - Posvećena karijeri
  - Nedostaje joj vremena za sebe
  - Uživa u Netflix-u

Demografski profil pruža samo “suve” činjenice o Meredith, dok se više detalja dobija psihografskom analizom. Kombinovanjem ovih informacija stvara se realna slika o Meredith kao potencijalnom klijentu, njenim potrebama, kupovnom potencijalu...

**Bihevioralna segmentacija** (eng. *Behavioral segmentation*) “deli” potrošače prema njihovom zapaženom ponašanju [9]. Marketinški stručnjaci smatraju da su bihevioralne varijable superiornije od geografskih i demografskih podataka pri kreiranju tržišnih segmenata. Neke od varijabli ponašanja su: prilika za kupovinu (specijalna prilika, poklon, praznici), status klijenta (stalni klijent, novi klijent, nije klijent), lojalnost prema brendu, spremnost kupca na kupovinu, i stav kupca. Dakle, nije bitno ko je kupac, već i kako se ponaša. Ova vrsta segmentacije omogućava kompanijama da preciznije usmere svoje marketinške aktivnosti (kampanje),

optimizuju komunikaciju sa potrošačima i poboljšaju korisničko iskustvo. Na primer, segmentisanjem korisnika prema učestalosti kupovine ili sklonosti ka određenim promocijama, kompanije mogu prilagoditi svoje ponude, čime se povećava verovatnoća uspešne realizacije prodaja. Bihevioralna segmentacija takođe pomaže u identifikaciji potencijalno profitabilnih kupaca, kao i onih koji su skloni odustajanju od kupovine, omogućavajući fokusiranje resursa na zadržavanju ključnih klijenata. Praćenjem ponašanja kupaca, kao što su navike otvaranja e-poruka ili vreme provedeno na sajtu, kompanije mogu prilagoditi svoje strategije kako bi poboljšale lojalnost kupaca i povećale prihod.

Kupci često traže različite “koristi” prilikom kupovine proizvoda. Na primer, u oblasti nege kože, ljudi imaju različite prioritete. Neki žele proizvode koji pomažu kod masne kože, neki pak traže proizvode koji uklanjaju tamne podočnjake, dok su neke interesuje ambalaža, miris, poreklo sastojaka itd. Sephora, na primer, u svojoj online prodavnici nudi opciju „Shop by Concern“ kako bi pratila specifične potrebe kupaca.

Popularni brendovi vrše segmentaciju kupaca prema njihovim interesovanjima kako bi pružili relevantne i personalizovane preporuke. Netflix i Spotify, na primer, koriste pametne preporuke prateći aktivnosti korisnika tj sadržaj kojem on pristupa (žanr filmova koji često gleda ili vrstu muzike koju često sluša). Na osnovu prikupljenih informacija, aplikacija će korisnicima nuditi više sličnog sadržaja, i time verovatno doprineti pozitivnijem korisničkom iskustvu.

Kupci razmatraju razne faktore pre donošenja odluke o kupovini. Razmenjuju mišljenja sa prijateljima, upoređuju cene u različitim prodavnicama, čitaju recenzije. Proces donošenja odluke o kupovini može biti složen. Ključni faktori koji utiču na njihovu odluku svakako uključuju percepciju brenda, akcijska ponude, sniženja i sl.

**Tehnografska segmentacija** (eng. *Technographic segmentation*) podrazumeva kategorizaciju kupaca na osnovu njihove upotrebe tehnologije, uključujući uređaje koje koriste, softverske preferencije i nivo prihvatanja novih tehnologija [9]. Na primer, kompanija može podeliti svoje kupce na one koji koriste pametne telefone, tablete ili desktop računare, i na osnovu toga prilagoditi svoje marketinške strategije. Razumevanjem tehnoloških preferencija svojih korisnika, kompanije mogu razviti proizvode koji su kompatibilni s njihovim potrebama i usmeriti reklame na prave platforme.

**Segmentacija zasnovana na vrednostima** (eng. *Value-based segmentation*) odnosi se na deljenje kupaca u grupe na osnovu ekonomske vrednosti koju donose kompaniji. Ova vrednost može uključivati koliko često kupac kupuje, koliko novca troši, i koliki je ukupni profit koji kompanija ostvaruje od tog kupca tokom vremena. Na primer, kompanija može prepoznati kupce koji troše više novca i zadržavaju se duže vreme kao kupce visoke vrednosti. Samim tim, kompanija ovoj grupi kupaca nudi posebne ponude, nagrade kroz programe lojalnosti i sl. kako bi ih podstakli da nastavu da kupuju; dok kupci koji donose manju vrednost mogu biti ciljani drugačijim strategijama, kao što su ponude koje će ih motivisati da povećaju svoju potrošnju.

**Segmentacija zasnovana na potrebama** (eng. *Needs-based segmentation*) fokusira se na razumevanje osnovnih razloga zbog kojih kupci kupuju određeni proizvod ili uslugu, te ih deli u grupe prema njihovim specifičnim potrebama i željama. Na primer, putnička agencija može

“podeliti” svoje korisnike na one koji traže luksuzna putovanja, avanturistička iskustva ili budžetske opcije. Ova vrsta segmentacije omogućava kompanijama da kreiraju prilagođene proizvode i usluge koje direktno zadovoljavaju specifične potrebe kupaca, čime povećavaju njihovo zadovoljstvo.

Postoji niz metoda za segmentaciju kupaca koje se razlikuju po svojoj kompleksnosti i pristupu. Tradicionalne metode uključuju RFM analizu, koja ocenjuje kupce na osnovu tri ključna faktora: koliko je nedavno kupac obavio kupovinu, koliko često kupuje, i koliko novca troši. Ovaj pristup pomaže u identifikaciji najvrednijih kupaca i CLV analizu, koja procenjuje ukupnu vrednost kupca za kompaniju i pokušava da, na osnovu njegovog prethodnog ponašanja, predvidi koliko će biti vredan kompaniji u budućnosti. Savremenije tehnike obuhvataju različite algoritme klasterizacije i neuronske mreže. Algoritmi klasterizacije grupišu kupce u segmente na osnovu sličnosti njihovih karakteristika. Popularni algoritmi obuhvataju K-Means algoritam, koji deli podatke u unapred definisani broj klastera, hijerarhijski klastering, koji gradi hijerarhijsku strukturu klastera, DBSCAN, koji identifikuje klastere na osnovu “gustine” podataka i može da pronade neobične zavisnosti itd. Neuronske mreže koriste složene modele za identifikaciju obrazaca u velikim i složenim skupovima podataka. Tehnike kao što su autoenkoderi za redukciju dimenzionalnosti i *deep-learning* modeli za složeno učenje i analizu omogućavaju napredniju segmentaciju.

### 2.3.2 RFM model

RFM (*Recency, Frequency, Monetary*) model predstavlja metodologiju koja se koristi za analizu i segmentaciju kupaca na osnovu njihove kupovne aktivnosti. Ovaj model pomaže marketinškim stručnjacima da lakše odrede ciljne grupe među kupcima i postave odgovarajuće marketinške strategije. Ovaj pristup uključuje procenu ponašanja kupaca na osnovu tri kriterijuma: skorašnjost kupovine (*R*), učestalost kupovine (*F*) i novčana vrednost kupovine (*M*) [11].

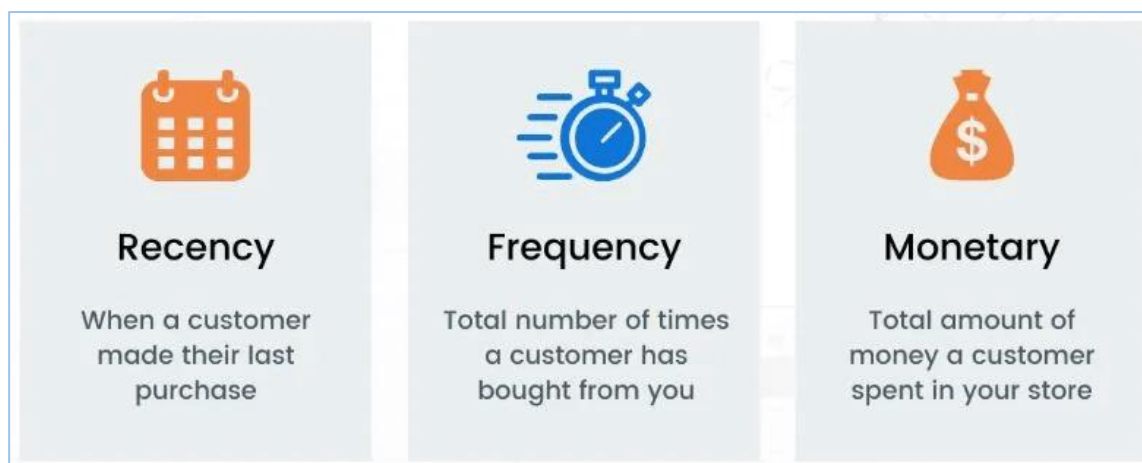
*Recency* (*R*) je mera koja pokazuje koliko je vremena prošlo od poslednje interakcije klijenta sa kompanijom. U većini slučajeva, ova aktivnost se odnosi na kupovinu, ali može uključivati i druge oblike interakcije, kao što su poslednja poseta veb sajtu ili korišćenje mobilne aplikacije. Što je skorije klijent imao neki oblik interakcije sa kompanijom, veća je verovatnoća da će biti “otvoreniji” za dalju komunikaciju i ponude od strane kompanije/brenda. Ovaj parameter ukazuje na to koliko je kupac aktivan i raspoložen za naredne marketinške ponude.

*Frequency* (*F*) predstavlja učestalost interakcije klijenta sa kompanijom ili kupca sa brendom u određenom vremenskom periodu. Ovaj faktor procenjuje koliko često kupac obavlja transakcije, pri čemu veća učestalost ukazuje na veću lojalnost i zadovoljstvo kupca. Kupci koji češće kupuju proizvode/usluge pokazuju veći stepen angažovanosti, što može značiti da su zadovoljniji ponuđenim proizvodima/uslugama. S druge strane, ne tako česte transakcije ukazuju na potrebu za dodatnim marketinškim ponudama kako bi se kupac ponovo aktivirao. Takođe,

kupci koji su obavili samo jednu kupovinu predstavljaju posebnu kategoriju kojoj je potrebno posvetiti dodatnu pažnju.

*Monetary (M)* predstavlja ukupan novčani iznos koji je kupac potrošio na kupovine u određenom vremenskom periodu. Klijenti koji troše više novca kupujući proizvode/usluge brenda značajnije doprinose prihodima kompanije. Samim tim, ovakvi kupci obično uživaju poseban tretman. Neretka je pojava da klijenti koji troše najviše novca ne kupuju često, ali kada to čine, biraju skuplje proizvode/usluge. Posmatranje monetarne vrednosti podeljene sa učestalošću kupovine pokazuje prosečan iznos jedne kupine – što je važan sekundarni faktor koji treba uzeti u obzir prilikom segmentacije kupaca.

**RFE** (*Recency, Frequency, Engagement*) model je varijacija RFM modela i koristi se kada monetarni aspekt nije relevantan, kao što je slučaj kod praćenja gledanosti, slušanosti, čitanosti ili interakcije sa sadržajima na internetu. Umesto monetarnog faktora, u RFE modelu se koristi parametar angažovanja, koji se definiše kao kombinovana vrednost zasnovana na različitim metrikama, poput broja poseta jednoj stranici, trajanja tih poseta, broja pregledanih stranica, i vremena provedenog na njima. Ovaj pristup omogućava kompanijama da procene lojalnost i interesovanje korisnika na osnovu njihovog ponašanja, što je naročito korisno u oblastima gde finansijske transakcije nisu primarni pokazatelj angažovanja.



**Slika 3.** RFM model

Formiranje RFM modela obuhvata nekoliko ključnih koraka sa ciljem segmentacije kupaca na osnovu njihove kupovne aktivnosti. Prvi korak podrazumeva prikupljanje podataka o kupovinama u određenom vremenskom periodu, uključujući datum svake transakcije, broj obavljenih transakcija i iznos potrošnje. Ovi podaci direktno utiču na izračunavanje RFM vrednosti.

Sledeći korak je izračunavanje RFM vrednosti – procenjuje se koliko je vremena prošlo od poslednje kupovine ili interakcije klijenta sa brendom (R), učestalost kupovina u tom periodu (F), i ukupna količina novca koju je kupac/klijent potrošio (M).

Customer ID	Recency [Day]	Frequency [Number]	Monetary [Total]
1	4	6	540
2	6	11	940
3	64	1	35
4	23	3	65
5	15	4	179
6	32	2	56
7	7	3	140
8	50	1	950
9	34	15	2630
10	10	5	191
11	3	8	845
12	1	10	1510
13	27	3	54
14	18	2	40
15	5	1	25

**Tabela 2.** Primer skupa podataka o transakcijama kupaca

Zatim se vrši rangiranje klijenata prema svakom od ovih atributa pojedinačno. Obično se koristi skala od 1 do 5, pri čemu 5 predstavlja najbolje rezultate (npr. najskorija kupovina, najveća učestalost kupovine, najveći iznos potrošnje).

Customer ID	Recency	R Score	Customer ID	Frequency	F Score	Customer ID	Monetary	M Score
12	1	5	9	15	5	9	2630	5
11	3	5	2	11	5	12	1510	5
1	4	5	12	10	5	8	950	5
15	5	4	11	8	4	2	940	4
2	6	4	1	6	4	11	845	4
7	7	4	10	5	4	1	540	4
10	10	3	5	4	3	10	191	3
5	15	3	13	3	3	5	179	3
14	18	3	7	3	3	7	140	3
4	23	2	4	3	2	4	65	2
13	27	2	14	2	2	6	56	2
6	32	2	6	2	2	13	54	2
9	34	1	15	1	1	14	40	1
3	46	1	8	1	1	3	35	1
8	50	1	3	1	1	15	25	1

**Tabela 3.** Primer rangiranja kupaca po svakom od atributa R, M, F pojedinačno

Naredni korak je kombinovanje individualnih R, F, M rezultata i izračunavanje finalnog RFM rezultata. RFM rezultat prikazan u tabeli u nastavku je izračunat kao prosečna vrednost

pojedinačnih rezultata po kategorijama. U ovom primeru je svakom data podjednaka važnost. U zavisnosti od prirode biznisa, svaka od ovih vrednosti može da učestvuje sa određenom „težinom“ tj. Da manje ili više učestvuje u objedinjenom RFM rezultatu. Na primer, kada je u pitanju kupovina proizvoda kao što su televizor, frižider i sl. Monetarna vrednost po transakciji je obično visoka, ali učestalost kupovine i skorašnjost imaju niže vrednosti. U takvim situacijama parametar učestalosti bi trebalo da ima najmanju težinu tj da najmanje utiče na ukupan rezultat. U maloprodaji koja se bavi modom ili kozmetikom, redovni kupci koji obavljaju transakcije bar jednom mesečno imaju veći R I F rezultat, pa bi trebalo da oni učestvuju s većim težinama u konačnom rezultatu.

Customer ID	RFM	RFM Score
1	5,4,4	4.3
2	4,5,4	4.3
3	1,1,1	1.0
4	2,2,2	2.0
5	3,3,3	3.0
6	2,2,2	2.0
7	4,3,3	3.3
8	1,1,5	2.3
9	1,5,5	3.7
10	3,4,4	3.3
11	5,4,4	4.3
12	5,5,5	5.0
13	2,3,2	2.3
14	3,2,1	2.0
15	4,1,1	2.0

**Tabela 4.** Primer konačnog RFM rezultata

U narednom koraku se vrši klasifikacija kupaca u klase kupaca sa sličnim kupovnim karakteristikama. U svakoj od ovih klasa detaljno se analiziraju obrasci ponašanja kupaca, čime se identifikuju specifične potrebe i preferencije. Na osnovu rezultata RFM analize, razvijaju se marketinške strategije prilagođene različitim klasama kupaca, poput posebnih ponuda za najlojalnije kupce ili strategija za reaktiviranje manje aktivnih kupaca. Efikasnost ovih strategija se redovno prati i analiza se ažurira kako bi se uvela nova rešenja u skladu sa promenama u ponašanju kupaca.

Često se u odnosu na krajnji rezultat RFM analize vrši podela u 11 klasa. U nastavku je prikazana tabela koja sadrži opis svake klase, zajedno sa preporukama za dalje marketinške akcije usmerene na svaku od njih [12].



<i>Kategorija</i>	<i>Aktivnost</i>	<i>Preporuka</i>
<i>Champion</i>	Nedavno kupovali. Često kupuju. Troše veliku svotu novca.	Nagraditi ih. Mogu biti najraniji usvojioi novih proizvoda. Promovišu brend.
<i>Loyal</i>	Redovno kupuju. Koriste promocije.	Ubediti ih da kupe skuplje proizvode. Tražiti recenzije.
<i>Potential loyalist</i>	Skorašnji kupci koji troše pristojnu svotu novca.	Ponuditi im program lojalnosti. Ponuditi im lične preporuke
<i>New customer</i>	Najčešće kupuju	Pružiti im pomoć pri kupovini, razvijati lojalniji odnos sa njima.
<i>Promising</i>	Često kupuju i troše pristojnu svotu novca ali su poslednju transakciju obavili pre više nedelja.	Ponuditi im kupone. Vratiti ih na platformu i pobrinuti se da budu angažovani. Ponuditi im lične preporuke.
<i>Core</i>	Standardni kupci sa skorašnjom poslednjom kupovinom.	Ponuditi im promocije.
<i>Needs attention</i>	Stari klijenti koji su poslednju transakciju obavili pre više od mesec dana.	Ponuditi im promocije i lične preporuke.
<i>Can't lose them but losing</i>	Obavljaju najveće transakcije i to rade često ali nisu bili aktivni duže vreme.	Pridobiti ih ponovo novim proizvodima. Ne dozvoliti da ih privuče konkurencija.
<i>At risk</i>	Slični prethodnim, ali to čine ređe i troše manje svote novca.	Slati im mejlove, obezbediti resurse na sajtu koji će im koristiti.
<i>Losing but engaged</i>	Poslednju kupovinu su obavili odavno ali redovno posećuju sajt i proveravaju email.	Probuditi njihovo interesovanje popustima na određene proizvode. Slati personalizovane mejlove.
<i>Lost</i>	Poslednju kupovinu su obavili odavno i dalje se nisu angažovali.	Probuditi interesovanje kampanjama koje imaju zadatak da dopru do klijenata. Ignorirati sve ostalo

**Tabela 5.** RFM kategorije klijenata

*Champion* – Ovi kupci predstavljaju ključne aktere koji donose veliki deo prihoda. Neophodno je uložiti dodatne napore da bi ostali zadovoljni, uključujući ekskluzivne ponude i popuste. Njihova vrednost za kompaniju se može dodatno povećati njihovim promovisanjem u brend ambasadore.

*Loyal* – Ovi kupci su aktivni i donose značajan prihod. Već postoji dobra komunikacija s njima, ali cilj je dodatno poboljšati njihovo iskustvo i lojalnost. Ne treba im slati neželjenu e-poštu, već preporuke za manji broj proizvoda koji su njima interesantni. Poželjno bi bilo zatražiti od njih da podelje svoja pozitivna iskustva sa ostalim potrošačima.

*Potential Loyalist* – Ovi kupci su već nekoliko puta obavili kupovinu, ali njihove transakcije nisu bile tako velike. Da bi se to promenilo, potrebno je motivisati ih putem preporuka za unakrsnu prodaju i promocija koje podstiču veće kupovine.

*New customer* – Ovi kupci su relativno nedavno obavili kupovinu i njihova potrošnja je prosečna ili ispod proseka. Imaju potencijal da postanu vredni kupci. Preporučuje se ponuda popusta na dodatne proizvode i uključivanje u redovnu marketinšku komunikaciju, kao i slanje anketa za bolje razumevanje njihovih interesovanja.

*Promising* – Ovi kupci su relativno nedavno obavili veliku kupovinu i doneli značajniji prihod, ali još uvek nisu postali redovni klijenti. Neophodno je uložiti u njihovu lojalnost kroz posebne ponude i informacije o dodatnim proizvodima/uslugama koji su za njih relevantni.

*Need attention* – Ovi kupci su u fazi razmišljanja o sledećoj kupovini i postoji opasnost da ih privuče konkurencija. Zato je potrebno dodatno ih motivisati i pridobiti ih na svoju stranu korišćenjem vremenski ograničenih promocija i popusta.

*At risk* – Ovi kupci su nedavno obavili kupovinu ali postoji opasnost da će prestati da kupuju i dalje, ukoliko se ne preduzmu mere. Preporučuje se personalizovani pristup koji uključuje preporuke na osnovu njihove prethodne aktivnosti i ponuda novih proizvoda/usluga u skladu sa njihovim interesovanjima, kako bi se povećala verovatnoća njihovog ponovnog angažovanja.

*Can't lose* – Ovi kupci su ranije bili lojalni ali su se nedavno okrenuli ka konkurenciji. Treba ih privući nazad poklonima, popustima i promocijama. Preporučuju se slične marketinške kampanje kao kod *at risk* kupaca. Dodatno, potrebno je utvrditi uzroke njihovog nezadovoljstva.

*Losing but engaged* – Ovi kupci su nedavno prestali da kupuju, ali su još uvek angažovani. Imajući u vidu njihovu prethodnu aktivnost, postoji šansa da se vrate ako im se ponudi atraktivan popust ili promocija. Oživljavanje interesa može biti ključno za njihov povratak.

*Lost* – Ovi kupci imaju mali uticaj na ukupni prihod kompanije i retko obavljaju kupovine. Njihova niska aktivnost i minimalna potrošnja čine ih najmanje prioritarnim za marketinške aktivnosti. Ulaganje resursa u ove kupce je obično neisplativo jer je verovatnoća njihovog povratka mala.

RFM model, sam po sebi, ne koristi mašinsko učenje, već se oslanja na informacije i pravila dobijenih iz podataka. Međutim, u savremenoj analitici i upravljanju odnosima sa kupcima, ML tehnike se često koriste za unapređenje RFM segmentacije. Korišćenjem algoritama mašinskog učenja, moguće je analizirati složenije obrasce i uključiti širi spektar varijabli, što omogućava sofisticiraniju i dinamičniju segmentaciju kupaca i optimizaciju marketinških strategija.

### **2.3.3 CLV model**

CLV (eng. *Customer Lifetime Value*) je model koji procenjuje ukupnu vrednost koju jedan kupac donosi kompaniji tokom trajanja njegovog odnosa sa brendom. CLV vrednost je važna jer je često “jeftinije” zadržati postojeće klijente nego privući nove. Fokusiranje na povećanje trenutne CLV vrednosti je dobar način ostvarivanje profita. Umesto da se kompanija

oslanja na pridobijanje novih klijenata (što može zahtevati više resursa), jednostavnije je razumeti šta trenutne kupce čini lojalnim i delovati u skladu sa tim. CLV je takođe odličan pokazatelj ranih znakova gubitka klijenata. Na primer, ako primetite da CLV opada i zaključite da kupci ne obnavljaju pretplatu na vaš proizvod ili uslugu, možete pokrenuti ili unaprediti program lojalnosti kako biste ih vratili. Takođe, poboljšanje korisničke podrške ili pojačavanje marketinga oko vremena obnavljanja pretplate može podstaći kupce da se ponovo prijave, čime se povećava CLV i prihod kompanije [13].

Istorijska CLV vrednost (eng. *Historic customer lifetime value*) – Ova vrednost pokazuje koliko je kupac do sada doprineo kompaniji. Na primer, ako ste poslednjih 10 godina svake godine kupovali božićno drvo od istog prodavca po ceni od 40 dolara, vaša ukupna CLV vrednost bi bila 400 dolara. Ova metoda pomaže da se shvati vrednost koju postojeći kupac donosi i pomaže u kreiranju profila idealnog kupca. Međutim, nije toliko korisna za predviđanje buduće vrednosti kupca..

Prediktivna CLV vrednost (eng. *Predictive customer lifetime value*) – Ova vrednost se izračunava pomoću algoritama koji koriste istorijske podatke kako bi predvideli koliko dugo će odnos sa kupcem trajati i koja će biti njegova buduća vrednost. Ovaj pristup uzima u obzir troškove pridobijanja kupaca, prosečnu učestalost kupovine, poslovne troškove i druge faktore kako bi dao realističniju prognozu. Iako je složeniji za izračunavanje, prediktivni CLV vam može pomoći da prepoznate kada je potrebno dodatno ulaganje u lojalnost kupaca.

Za izračunavanje CLV vrednosti neophodno je odrediti ključne indikatore performansi (KPIs) koji utiču na ovu metriku:

- Prosečna vrednost kupovine (eng. *Average Order Value, AOV*)
- Učestanost kupovine (eng. *Frequency, F*)
- Bruto marža (eng. *Gross Margin, GM*)
- Stopa odustajanja (eng. *Churn Rate, CR*)

Koristeći sve ove metrike, možemo primeniti sledeću formulu za izračunavanje CLV vrednosti:

$$CLV = AOV \times F \times GM \times \frac{1}{CR},$$
$$AOV = \frac{\text{Ukupan prihod od prodaje}}{\text{Ukupan broj kupovina}}$$

Na primer, ukoliko je ukupni prihod kompanije u određenom periodu 1 000 000 dolara, a ukupan broj kupovina 40 000, prosečna vrednost kupovine iznosi 25 dolara.

$$F = \frac{\text{Ukupan broj kupovina}}{\text{Ukupan broj jedinstvenih kupaca}}$$

Na primer, ukoliko je ukupan broj kupovina u određenom period 40 000 a broj jedinstvenih kupaca je 10 000, tada učestanost kupovine iznosi 4.

Bruto marža predstavlja procenat profita nakon što se oduzmu svi direktni troškovi proizvodnje ili nabavke robe ili usluga koje kompanija prodaje:

$$GM = \frac{\text{Ukupni prihod od prodaje} - \text{Troškovi prodaje}}{\text{Ukupni prihodi prodaje}}$$

Na primer, ukoliko su ukupni prihodi kompanije 800 000 dolara, a troškovi prodaje 470 000 dolara, tada bruto marža iznosi 41%.

$$CR = \frac{\text{Broj kupaca na kraju perioda} - \text{broj kupaca na početku perioda}}{\text{Broj kupaca na početku perioda}}$$

Neka je stopa odustajanja 50%; to znači da je period trajanja kupaca 2. Sledi da je ukupna CLV vrednost posmatrane kompanije:

$$CLV = 25 \times 4 \times 0.41 \times 2 = 82$$

To znači da je ukupna vrednost kupca tokom njegovog životnog veka 82 dolara.

**BGD/ND** (eng. *Beta-Geometric/Negative Binomial Distribution*) je statistički model koji se koristi za predviđanje ponašanja kupaca, određivanje njihove ukupne vrednosti i identifikovanje onih koji donose najveću vrednost kompaniji. Ovaj model koristi negativnu binomnu distribuciju da opiše broj kupovina koje kupac obavlja dok je aktivan. Svaki kupac ima specifičnu stopu kupovine. Beta-geometrijska distribucija procenjuje šanse da kupac prestane da bude aktivan u nekom trenutku. Svaki kupac ima različitu sklonost ka prestanku kupovine, ali i ova sklonost varira među kupcima. Ovaj model je popularan jer može efikasno da predvidi buduće ponašanje kupaca, što je posebno korisno u online trgovini. Takođe, može se koristiti čak i kada kompanija ne raspolaže sa dovoljno podataka (npr. nedostaju podaci o kupcima koji su relativno novi). Ipak, model ima određena ograničenja. Jedno od njih je pretpostavka da kupci prestaju da budu aktivni nasumično (*churn-rate*), što nije uvek slučaj budući da razni faktori mogu uticati na odluku kupaca da prestanu sa kupovinama [14].

**Gamma-Gamma** model je statistički alat koji se koristi u kombinaciji sa BG/NBD modelom kako bi se preciznije procenila ukupna vrednost kupca (CLV). Dok BG/NBD model predviđa koliko često će kupac obavljati kupovine, Gamma-Gamma model se fokusira na procenu prosečne vrednosti tih transakcija. Pretpostavka ovog modela je da ne postoji povezanost između učestalosti kupovine i vrednosti transakcija, što omogućava precizniju procenu prosečne potrošnje po transakciji. Na osnovu prošlih podataka o kupovinama, model može da predvidi prosečnu vrednost budućih kupovina. Kada se koristi zajedno sa BG/NBD modelom, Gamma-Gamma model pomaže da se dobije jasnija slika o ukupnoj vrednosti koju kupac donosi kompaniji tokom vremena. Međutim, važno je napomenuti da ovaj model pretpostavlja nezavisnost između učestalosti i vrednosti kupovina, što možda nije uvek slučaj u stvarnosti [14].

### 2.3.4 Metod klasterizacije

Klasterizacija (eng. *Clustering*) se odnosi na skup metoda za nenadgledano učenje (eng. *Unsupervised learning*) koje su zasnovane na statističkim i matematičkim principima. Ove metode su posebno korisne za otkrivanje skrivenih obrazaca i grupisanje podataka u složenim skupovima podataka. U kontekstu segmentacije kupaca, klasterizacija je od suštinskog značaja za identifikaciju i analizu različitih grupa potrošača prema njihovim sličnostima i razlikama [15].

Postupak segmentacije kupaca korišćenjem klasterizacije podrazumeva nekoliko ključnih koraka. Prvi korak je prikupljanje relevantnih podataka o kupcima, uključujući demografske informacije kao što su starost, pol i prihod, kao i podatke o kupovini poput učestalosti transakcija, vrednosti kupovina i “skorašnjosti”. Sledi pretprocesiranje podataka, koje uključuje čišćenje podataka kako bi se uklonile nepotpune ili netačne informacije, normalizaciju podataka kako bi se osigurala doslednost, i selektovanje ključnih karakteristika za klasterizaciju. Izbor algoritma klasterizacije zavisi od prirode podataka i ciljeva segmentacije. Na primer, ako se očekuje da klasteri budu slične veličine, može se koristiti K-Means, dok je DBSCAN prikladniji za klasterne nepravilnih oblika. Određivanje broja klastera, posebno kod K-Means algoritma, može se postići metodama poput Elbow metode ili Silhouette analize. Nakon klasterizacije, analizira se svaki klaster kako bi se razumeli njegovi atributi i ponašanje kupaca, koristeći različite tipove vizualizacija, kao što su scatter plotovi ili heatmap-e. Na osnovu ovih analiza, kompanije formiraju marketinške strategije, prilagođavaju proizvode i unapređuju korisničku podršku specifičnim segmentima kupaca.

**Klasterizacija zasnovana na centroidima** (eng. *Centroid-based Clustering*) je jedan od najjednostavnijih algoritama za klasterizaciju. Ovi algoritmi grupišu podatke na osnovu njihove blizine. Najčešće korišćene mere sličnosti su Euklidska distanca, Manhattan distanca ili Minkovski distanca. Podaci se razdvajaju u unapred određeni broj klastera, pri čemu svaki klaster predstavlja skup vrednosti. U poređenju sa ovim vrednostima, podaci se dodeljuju klasterima. Glavni nedostatak ovih algoritama je potreba da se unapred odredi broj klastera, što se može uraditi intuitivno ili pomoću metoda kao što je Elbow. Najpoznatiji predstavnik algoritama ove grupe je **KMeans**.

**Klasterizacija zasnovana na gustini** (eng. *Density-based Clustering*) identifikuje klasterne na osnovu gustine podataka. Za razliku od metoda zasnovanih na centroidima, koje zahtevaju unapred određivanje broja klastera i mogu biti osetljive na inicijalne uslove, metode zasnovane na gustini automatski određuju broj klastera i manje su osetljive na početne uslove. Ove metode su izuzetno efikasne za rad sa klasterima različitih veličina i oblika i mogu identifikovati klasterne u područjima sa različitim gustinama. Najpoznatiji algoritam ove vrste je **DBSCAN**.

**Klasterizacija zasnovana na povezanosti** (eng. *Connectivity-based Clustering*) raspoređuje podatke u hijerarhijski strukturirane klasterne. Počinje se sa svakim podatkom kao posebnim klasterom, koji se zatim spaja sa najbližijim klasterima sve dok se svi podaci ne

kombinuju u jedan veliki klaster. Ova metoda koristi dendrogram, stablo koje prikazuje kako se klasteri spajaju. Postoje dva pristupa:

- **Divizivna klasterizacija:** Inicijalno, svi podaci su smešteni u jedan klaster koji se zatim postepeno deli na manje klastere.
- **Agregativna klasterizacija:** Inicijalno, svi podaci su smešteni u pojedinačnim klasterima koji se zatim postepeno spajaju u veće klastere. Ovaj metod omogućava istraživanje različitih nivoa granularnosti klastera, pri čemu se dendrogram može preseći na određenoj visini kako bi se dobio željeni broj klastera.

**Klasterizacija zasnovana na distribuciji** (eng. *Distribution-based Clustering*) grupiše podatke prema njihovoj verovatnoći da pripadaju istoj statističkoj distribuciji, kao što je Gausova distribucija. Ova metoda koristi verovatnoće da bi grupisala podatke i dodeljuje podatke klasterima u skladu sa njihovom verovatnoćom pripadnosti. Glavni nedostatak ovih metoda je potreba za unapred određenim oblikom klastera, što može biti izazovno prilagoditi i može dovesti do prekomernog prilagođavanja (overfitting). Najpoznatiji algoritam ove vrste je **Gaussian Mixture Model (GMM)**, koji nudi fleksibilnost i preciznost u modelovanju klastera sa različitim distribucijama.

### 2.3.5 Neuronske mreže

Neuronske mreže su se pokazale efikasnim alatom za segmentaciju kupaca, naročito kada se radi o velikim i složenim skupovima podataka, gde tradicionalne metode klasterizacije, kao što je K-means, mogu imati potreškoća. Ključna prednost neuronskih mreža u odnosu na ostale pristupe leži u njihovoj sposobnosti da prepoznaju nelinearne obrasce u podacima, i da obrađuju podatke visoke dimenzionalnosti, što ih čini pogodnim za dublje uvide u segmentaciju kupaca. Neki od uobičajnijih prisupa su:

- Autoenkoderi - nenadgledane neuronske mreže (eng. *Unsupervised neural networks*) koje se koriste za učenje iz podataka bez eksplicitnih labela (kao što su etikete za klasifikaciju), već pokušavaju da “nauče” skrivena pravila ili obrasce iz podataka. Autoenkoderi uzimaju ulazne podatke (kao što su podaci o kupcima) i propuštaju ih kroz neuronsku mrežu koja je dizajnirana da nauči sažetu reprezentaciju tih podataka, tj. suštinu. Ovaj proces smanjuje dimenzionalnost podataka, što znači da autoenkoder pokušava da sažme važne informacije o svakom kupcu u manji broj karakteristika. Sažete reprezentacije podataka se koriste za grupisanje kupaca sličnih karakteristika u odgovarajuće segmente.
- Samoorganizujuće mape (eng. *Self-Organizing Maps - SOM*) - neuronske mreže koje koriste visoko dimenzione podatke (npr. podaci o karakteristikama kupaca) i transformišu ih u nižedimenzioni prostor, obično u obliku dvodimenzionalne mreže ili rešetke. Kroz proces treniranja, SOM mreža uči da grupiše slične podatke (kupce sa sličnim karakteristikama) u susedne čvorove na mapi. To znači da kupci koji su najbliži jedni drugima završe bliže jedni drugima na mapi, što olakšava vizualizaciju i analizu grupa kupaca. SOM mreža

omogućava da vizuelno prepoznamo koje grupe kupaca imaju slična ponašanja ili osobine na osnovu načina na koji su raspoređeni na mapi.

- Modeli dubokog učenja (eng. *Deep Learning Models*) - Konvolucione neuronske mreže (CNN) i rekurentne neuronske mreže (RNN), predstavljaju napredne metode za analizu i razumevanje kompleksnih podataka, kao što je ponašanje kupaca. Oni su korisni zato što mogu raditi sa složenim, nestrukturiranim podacima poput teksta, slika ili vremenskih serija. Konvolucione neuronske mreže su posebno korisne za analizu slika i podataka sa prostornom strukturom. Iako su najpoznatije po primeni u oblasti obrade slika, mogu se koristiti i u drugim kontekstima, uključujući analizu nestrukturiranih podataka u svetu poslovanja. CNN koristi specijalne slojeve koji primenjuju filtere na ulazne podatke (npr. slike) kako bi prepoznale važne obrasce. U kontekstu ponašanja kupaca, ako imamo podatke u formi slika (npr. proizvodi koje kupci pregledavaju), CNN može analizirati te slike i povezati vizuelne karakteristike sa obrascima kupovine. Na primer, može analizirati koje vrste proizvoda privlače pažnju određenih kupaca, što može pomoći u segmentaciji kupaca na osnovu vizuelnih preferencija. CNN se takođe može koristiti za analizu tekstualnih podataka, kao što su recenzije kupaca. Filteri mogu otkriti ključne reči ili fraze u recenzijama koje ukazuju na pozitivne ili negativne reakcije kupaca, što omogućava bolje razumevanje njihovih stavova i osećanja prema proizvodima. Rekurentne neuronske mreže su dizajnirane za rad sa sekvencijalnim podacima, što ih čini idealnim za analizu vremenskih serija, ponašanja kupaca tokom vremena, ili podataka koji zavise od redosleda (sekvence) događaja. Za razliku od drugih modela, RNN može zapamtiti prethodne informacije u sekvenci i koristiti ih pri donošenju odluka o budućim informacijama. To znači da RNN ne gleda samo na trenutni podatak, već uzima u obzir i ono što se desilo ranije.. U kontekstu ponašanja kupaca, RNN je posebno koristan kada analiziramo vremenske serije podataka, kao što su nizovi aktivnosti koje kupac obavlja (npr. klikovi na sajt, pregledanje proizvoda, kupovine). Zamislmo da imamo podatke o tome kako se kupac ponašao na veb-sajtu tokom vremena (npr. koji proizvodi su pregledani, koje stranice su posetili). RNN može analizirati redosled tih interakcija i na osnovu toga predvideti šta bi kupac mogao sledeće da uradi, npr. da li će napustiti sajt bez kupovine ili izvršiti kupovinu [18].

Segmentacija kupaca i analiza potrošačke korpe su tehnike koje zajedno omogućavaju dublje razumevanje ponašanja potrošača, kao i definisanje i optimizaciju odgovarajućih poslovnih strategija. Segmentacija kupaca grupiše potrošače prema parametrima kao što su skorašnjost i učestanost kupovine, količina potrošenog novca i očekivana profitabilnost, čime kompanije mogu identifikovati najvrednije kupce ili one koji imaju potencijal da to postanu. Analiza potrošačke korpe se nadovezuje na segmentaciju time što istražuje koje proizvode kupci iz određenih segmenata najčešće kupuju zajedno. Ovo omogućava kompanijama da prilagode ponudu za svaki segment i razviju strategije za promocije ili prodajne podsticaje koji su specifično usmereni na obrasce kupovine unutar svakog segmenta. Na taj način, segmentacija definiše grupe kupaca, a analiza potrošačkih korpi otkriva detalje njihovog ponašanja, čime zajedno doprinose preciznijem targetiranju i povećanju ukupne vrednosti kupaca.

## 2.4 Analiza potrošačke korpe

Analiza potrošačke korpe (eng. *Market Basket Analysis, MBA*) je tehnika otkrivanja znanja koja se koristi za otkrivanje i razumevanje obrazaca kupovine u maloprodaji. *MBA* je proces koji traži veze između objekata koji „idu zajedno“ u nekom poslovnom kontekstu. Ova tehnika omogućava identifikaciju veza između proizvoda koje potrošači često kupuju zajedno, pružajući uvid u njihovo ponašanje i preferencije. Cilj analize potrošačke korpe je ne samo da se identifikuju kombinacije proizvoda koje se često javljaju zajedno u transakcijama, već i da se razumeju uzroci i posledice tih obrazaca [16]. Na primer, ukoliko *MBA* pokaže da kupci koji kupuju mleko često kupuju i hleb, to može poslužiti kao osnova za strateško pozicioniranje ovih proizvoda na policama, unakrsnu prodaju, ili ciljane marketinške kampanje.

U kontekstu *BI*, *MBA* ima značajan uticaj na donošenje odluka koje direktno utiču na povećanje profitabilnosti. Analizom podataka iz transakcija maloprodajna preduzeća optimizuju prostor u prodavnicama, razvijaju personalizovane preporuke za kupce, te kreiraju promocije koje podstiču dodatnu potrošnju. Na ovaj način, preduzeća ne samo da bolje razumeju svoje kupce, već i aktivno utiču na njihove kupovne odluke, što vodi ka većim prihodima i efikasnijem poslovanju.

Razlikuju se tri tipa analize potrošačke korpe:

- **Deskriptivna analiza potrošačke korpe** (eng. *Descriptive MBA*) – koristi se za identifikaciju obrazaca kupovine na osnovu podataka o prethodno obavljenim transakcijama, bez pravljenja predikcija o budućem ponašanju kupaca. Glavni cilj ove analize je otkrivanje asocijacija ili veza između različitih proizvoda koji se često kupuju zajedno. Na primer, analiza može otkriti da kupci koji kupe kafu često kupuju i šećer, pa se ovi proizvodi često pojavljuju zajedno u transakcijama.
- **Prediktivna analiza potrošačke korpe** (eng. *Predictive MBA*) – ova tehnika koristi metode *nadgledanog učenja* (eng. *Supervised learning*), kao što su regresija (eng. *regression*) i klasifikacija (eng. *classification*) za predviđanje ponašanja kupaca na osnovu postojećih obrazaca kupovine. Prediktivna analiza može pomoći u formiranju personalizovanih preporuka proizvoda, i ciljanih promocija usmerenih ka specifičnim kupcima. Takođe, ovaj model je koristan za predviđanje buduće potražnje za određenim proizvodima, a analizom atipičnih ili neuobičajenih obrazaca kupovine može prepoznati i potencijalne prevare.
- **Uporedna analiza potrošačke korpe** (eng. *Differential MBA*) – koristi se za poređenje obrazaca kupovine među različitim grupama potrošača. Umesto da se fokusira samo na određivanje artikala koji se često kupuju zajedno, ova analiza istražuje kako se ti obrasci razlikuju između različitih kupaca ili u različitim situacijama. Takođe, može se primeniti za poređenje kupovnih navika različitih društvenih grupa, sezonskih promena ili čak za poređenje ponašanja kupaca u odnosu na konkurenciju.



U analizi potrošačke korpe se obično koriste podaci iz kupoprodajnih transakcija, pri čemu je svaka transakcija predstavljena listom kupljenih artikala. U tabeli 6 je prikazan primer skupa transakcija.

<i>TID</i>	<i>Stavke</i>
1	{cookies, chocolate, ice cream, cake}
2	{cookies, chocolate, donut}
3	{cookies, chocolate, cake}
4	{bread, eggs, milk}
5	{pizza, sandwich}
6	{fries, sandwich}

**Tabela 6.** Skup transakcija za *MBA*

Nad ovakvim skupom podataka se primenjuje asocijativna analiza (eng. *Association Analysis*) kako bi se pronašle interesantne korelacije između artikala. Asocijativna pravila imaju oblik *ako-onda* tj.  $\{IF\} \rightarrow \{THEN\}$ . Ova pravila se koriste za predviđanje verovatnoće da će se određeni artikli naći u okviru iste transakcije.

U analizi potrošačke korpe često se koriste sledeći termini:

- Skup stavki (eng. *Itemset*) – lista artikala koji su kupljeni u istoj transakciji
- Podrška, potpora (eng. *Support*) – učestalost sa kojom se određeni skup artikala pojavljuje u bazi transakcija. Takođe se može izraziti kao verovatnoća. Na primer, ako se mleko pojavljuje u 50 od ukupno 500 transakcija, tada je verovatnoća 50/500 ili 0.1
- Pouzdanost (eng. *Confidence*) – verovatnoća da će se neki podskup artikala naći u istoj transakciji
- Uzrok (eng. *Antecedent*) – leva strana pravila (IF), podskup artikala iz skupa podataka
- Posledica (eng. *Consequent*) – desna strana pravila (THEN), podskup artikala koji je u relaciji sa antecedentom
- *Lift* – uzima u obzir statističke zavisnosti, odnos pouzdanosti pravila i potpore (verovatnoća pojavljivanja podskupa artikala s leve strane pravila). Ako je dobijena vrednost  $>1$  pravilo je pozitivno pridruženo, vrednost  $<1$  znači da je pravilo negativno pridruženo, dok 1 označava da je pravilo beznačajno.

Zadavanjem minimalne vrednosti potpore i pouzdanosti koje pravilo treba da zadovolji vrši se prihvatanje/odbacivanje pravila. Pravila sa malom potporom obično nisu od velike važnosti, tako ih možemo zanemariti.

Primer:  $\{cookies, chocolate\} \rightarrow \{cake\}$

- *lift*  $>1$  – kupovina kolačića i čokolade podrazumeva i kupovinu torte.
- *lift*  $<1$  – kupovina kolačića i čokolade verovatno ne znači i kupovinu torte.

### 2.4.1 Algoritmi za analizu potrošačke korpe

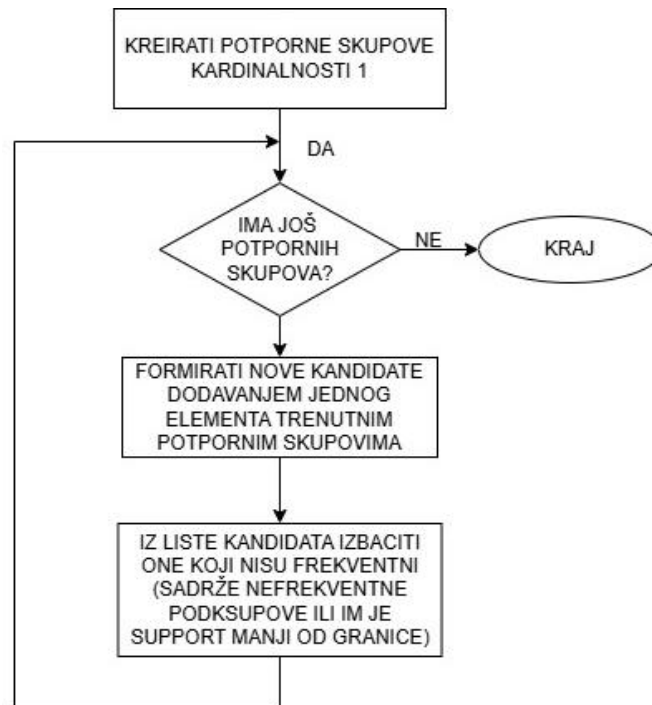
Ukoliko bi se asocijativna pravila formirala i proveravala metodom grube sile (eng. *brute force*), to bi oduzimalo previše vremena. Danas se za nalaženje asocijativnih pravila koriste različite *data mining* tehnike i algoritmi. Jedan od važnijih ciljeva prilikom predviđanja verovatnoće da će se određeni artikli naći zajedno u istoj transakciji je postizanje zadovoljavajuće tačnosti (eng. *accuracy*). U praksi, u analizi potrošačke korpe najčešće se koriste sledeći algoritmi:

- **Apriori**
- **FP Growth**
- **ECLAT**

Istraživanja su pokazala da klasični algoritmi poput Apriori i FP-Growth, koji koriste horizontalni pristup, zahtevaju više vremena za formiranje asocijativnih pravila, budući da se kroz bazu podataka “prolazi” više puta. ECLAT koristi vertikalni pristup koji eliminiše potrebu za višestrukim iteriranjem kroz bazu podataka, što značajno smanjuje vreme izvršenja. Ipak, vreme izvršenja nije jedini faktor koji treba uzeti u obzir pri odabiru algoritma. Apriori i FP Growth se i dalje aktivno koriste u analizi potrošačke korpe budući da daju dobre rezultate. FP Growth je prikladniji za velike skupove podataka, dok je Apriori dovoljno efikasan za manje i srednje skupove podataka, koji se obično i koriste u edukativne svrhe. U praktičnom delu rada će biti iskorišćen Apriori i FP Growth algoritmi.

#### Apriori algoritam

*Apriori* algoritam je jedan od najpoznatijih algoritama za učenje asocijativnih pravila i često se koristi za analizu podataka u relacionim bazama podataka. Cilj *Apriori* algoritma je identifikovanje stavki koje se često pojavljuju u istim transakcijama, i da što više smanji broj pravila u skupu kandidata. Bazira se na principu kreiranja skupova frekventnih pravila. Skup je frekventan ukoliko sadrži stavke čija je potpora (*support*) veća od neke unapred definisane vrednosti. Ako neki skup stavki ima potporu nižu od zadate vrednosti, taj skup se odbacuje. Ovo svojstvo zapravo kaže da ukoliko je neki skup stavki čest (frekventan), tada su i svi njegovi podskupovi frekventni. Time se značajno smanjuje broj skupova kandidata [16]. Na primer, ukoliko  $\{milk\}$  nije frekventan skup, tada nijedan širi skup koji sadrži mleko nije frekventan. Ova osobina se naziva antimonotonost.



**Slika 4.** *Apriori* algoritam. Kreiranje frekventnih skupova.

Na osnovu dobijenih frekventnih skupova potrebno je formirati asocijativna pravila. Asocijativna pravila su u osnovi implikacije koje ukazuju na odnose među stavkama u transakcijama. Ako je  $X$  posmatrani frekventni skup stavki, taj skup se “deli” na dva neprazna podskupa  $Y$  i  $X - Y$ , tako da pravilo  $Y \rightarrow X - Y$  ima pouzdanost  $c$  veću od neke unapred zadate vrednosti.

U opštem slučaju, pouzdanost nema osovину antimonotonosti, ali ukoliko su pravila izvedena iz istog skupa ( $X$ ) tada važi:

$$L = \{A, B, C, D\} \rightarrow c(ABC \rightarrow D) \rightarrow c(AB \rightarrow CD) \rightarrow c(A \rightarrow BCD)$$

Zahvaljujući ovom pravilu, odstranjuju se pravila koja ne ispunjavaju uslov minilane pouzdanosti. Naredni korak je spajanje onih asocijativnih pravila koja imaju isti prefiks sa posledične (desne) strane. Ukoliko novodobijeno pravilo nema zadovoljavajuću pouzdanost, odbacuje se. Na ovaj način samo pravila koja su dovoljno pouzdana i korisna ostaju u analizi.

$$(CD \rightarrow AB), (BD \rightarrow AC) \gg (D \rightarrow ABC)$$

### FP-Growth algoritam

FP-Growth (eng. *Frequent Pattern Growth*) je efikasan algoritam za otkrivanje često ponavljanih obrazaca u velikim skupovima podataka. Razvijen je kao alternativa *apriori* algoritmu, koji može biti spor i neefikasan za velike skupove podataka zbog učestalog skeniranja i generisanja frekventnih skupova kandidata. FP-Growth koristi strukturu nazvanu FP stablo (eng. *Frequent Pattern Tree*) koja predstavlja kompresovanu verziju baze tj. skupa podataka i omogućava efikasno pretraživanje i generisanje obrazaca. FP stablo se formira u dva koraka:

- Analiziraju se podaci kako bi se identifikovali artikli koji se često pojavljuju u transakcijama – obrasci sa jednim artiklom. Svaki artikl se dodaje u stablo zajedno sa svojom frekvencijom pojavljivanja. Ako transakcija deli deo sa već postojećom granom, stablo se gradi na način da se koristi već postojeća grana, dok se broj pojavljivanja proizvoda u toj grani povećava.
- Na osnovu prethodno formiranog FP stabla identifikuju se često pojavljivani obrasci dodavanjem proizvoda u grane stabla, stvarajući sve duže grane koje predstavljaju sve veće skupove proizvoda. Algoritam prvo identifikuje osnovne (minimalne) često pojavljujuće obrasce i onda dodaje dodatne proizvode jedan po jedan, proširujući grane stabla. Na taj način, algoritam generiše sve moguće kombinacije proizvoda koje čine često pojavljujuće skupove. To je tehnika rekurzivnog proširivanja.

Glavna prednost FP-Growth u odnosu na Apriori je brzina – FP-Growth je brži, posebno kada se radi sa velikim bazama podataka zahvaljujući smanjenom broju skeniranja baze podataka i eliminacije potrebe za formiranjem frekventnih skupova. Međutim, FP-Growth može postati memorijski izuzetno intenzivan kada FP stablo dostigne veliki obim, što može predstavljati izazov kod velikih skupova podataka [16].

## ECLAT

ECLAT (eng. *Equivalence Class Clustering and Association Rule Mining*) je još jedan algoritam za otkrivanje često ponavljanih obrazaca. Radi po principu prebrojavanja koliko puta se određeni artikli pojavljuju zajedno u transakcijama. ECLAT koristi matricu transakcija. U toj matrici, svaka vrsta predstavlja jednu transakciju, dok su po kolonama raspoređeni artikli. Čelije u matrici sadrže oznake koje ukazuju na prisutnost određenog artikla u određenoj transakciji. Algoritam pronalazi skupove proizvoda koji se često kupuju zajedno tako što prelazi preko svih redova u matrici i računa koliko puta se određeni proizvodi pojavljuju zajedno. Koristi presek (eng. *intersection*) skupova transakcija da identifikuje ove česte skupove. Kada pronađe skupove proizvoda, ECLAT izračunava koliko puta se svaki skup pojavio u bazi podataka. Na osnovu tih skupova, algoritam generiše asocijativna pravila. ECLAT koristi vertikalni pristup koji eliminiše potrebu za višestrukim iteriranjem kroz bazu podataka, što značajno smanjuje vreme izvršenja [16].

## Neuronske mreže

Veštačke neuronske mreže nalaze primenu kako u segmentaciji kupaca, tako i u analizi potrošačkih korpi. Korišćenje neuronskih mreža u analizi potrošačke korpe može značajno unaprediti uvid u podatke o kupovinama, posebno kada su setovi podataka veliki i složeni. Dok se tradicionalna analiza potrošačke korpe fokusira na statističke podatke o transakcijama, rekurentne neuronske mreže mogu da obrađuju sekvencijalne podatke, što je ključno kada je redosled kupovina bitan. Na primer, RNN može da predvidi sledeći proizvod koji će kupac verovatno kupiti na osnovu istorije njegovih kupovina, omogućavajući kompanijama da optimizuju sisteme za preporuke proizvoda i strategije unakrsne prodaje. Zamislite kupca koji prvo kupi hleb, zatim mleko, potom jaja. Tradicionalna analiza potrošačke korpe tretira ove proizvode kao statičan skup. Međutim, pomoću RNN možemo uočiti da je redosled kupovina

takođe važan - kupci koji kupe hleb često slede sa kupovinom mleka i jaja. RNN omogućava bolje prepoznavanje ovih sekvencijalnih obrazaca kupovine. Iako se konvolucione neuronske mreže obično koriste za analizu slika, mogu se adaptirati i za podatke o transakcijama. Kada se primene u analizi potrošačke korpe, CNN mreže mogu prepoznati lokalne obrasce unutar matrice transakcija. To znači da CNN može pomoći u otkrivanju proizvoda koji se često kupuju zajedno unutar podgrupa kupaca, zasnovano na sličnostima u njihovom ponašanju prilikom kupovine. Neuronske mreže se koriste i za kreiranje kolaborativnih filtracionih modela (eng. *Neural Collaborative Filtering, NCF*) koji preporučuju proizvode kupcima. Umesto da pronalaze proste asocijacije, NCF uči gustu, nelinearnu interakciju između kupaca i proizvoda. To omogućava modelu da bolje predvidi koje će proizvode kupci kupovati zajedno, pružajući personalizovanije preporuke [19].

### 3. Detalji implementacije

U praktičnom delu rada implementirane su segmentacija klijenata i analiza potrošačke korpe. U implementaciji je korišćen popularni *online retail* skup podataka preuzet sa adrese: <https://archive.ics.uci.edu/dataset/352/online+retail>, programski jezik *Python* i *Microsoft PowerBI* alat za vizuelizaciju dobijenih rezultata.

Korišćeni skup podataka sadrži sledeće atribute:

- CustomerID – jedinstveni identifikator kupca
- InvoiceNo – jedinstveni identifikator fakture; jedna faktura može sadržati više transakcija
- StockCode – jedinstveni identifikator skupa stavki u okviru fakture
- Description – lista kupljenih artikala u jednoj transakciji
- Quantity – količina, broj kupljenih artikala
- UnitPrice – jedinična cena artikla
- InvoiceDate – datum izvršenja transakcije
- Country – država u kojoj je izvršena transakcija

Na slici 5 prikazano je prvih nekoliko „ulaza“ iz skupa podataka.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850.0	United Kingdom
3	536365	840296	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/10 8:26	3.39	17850.0	United Kingdom

Slika 5. Primer podataka iz dataset-a

Posmatranjem skupa podataka, lako se uočava da atributi (kolone) Quantity i UnitPrice mogu imati i negativne vrednosti. To znači da u skupu podataka postoje transakcije koje predstavljaju povraćaj prethodno kupljenih artikala. Ovakve transakcije nisu od interesa, pa ih je potrebno ukloniti iz skupa podataka. Takođe, u nekim transakcijama nedostaju vrednosti CustomerID i/ili Description, pa i njih treba ukloniti iz skupa podataka.

Data columns (total 8 columns):				Data columns (total 8 columns):			
#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype
0	InvoiceNo	541909 non-null	object	0	InvoiceNo	397884 non-null	object
1	StockCode	541909 non-null	object	1	StockCode	397884 non-null	object
2	Description	540455 non-null	object	2	Description	397884 non-null	object
3	Quantity	541909 non-null	int64	3	Quantity	397884 non-null	int64
4	InvoiceDate	541909 non-null	object	4	InvoiceDate	397884 non-null	object
5	UnitPrice	541909 non-null	float64	5	UnitPrice	397884 non-null	float64
6	CustomerID	406829 non-null	float64	6	CustomerID	397884 non-null	float64
7	Country	541909 non-null	object	7	Country	397884 non-null	object
dtypes: float64(2), int64(1), object(5)				dtypes: float64(2), int64(1), object(5)			
memory usage: 33.1+ MB				memory usage: 27.3+ MB			
None				None			
(541909, 8)				(397884, 8)			

**Slika 6.** Osnovne informacije o skupu podataka pre i posle pretprocesiranja

	Quantity	UnitPrice	CustomerID
count	397884.000000	397884.000000	397884.000000
mean	12.988238	3.116488	15294.423453
std	179.331775	22.097877	1713.141560
min	1.000000	0.001000	12346.000000
25%	2.000000	1.250000	13969.000000
50%	6.000000	1.950000	15159.000000
75%	12.000000	3.750000	16795.000000
max	80995.000000	8142.750000	18287.000000

**Slika 7.** Statistički opis numeričkih vrednosti u prečišćenom skupu podataka

Nakon pretprocesiranja podataka potrebno je kreirati *RFM* model. Najpre treba izračunati vrednosti pojedinačnih parametara tj. proceniti skorašnjost (*R*), učestanost (*F*) i monetarnu vrednost (*M*). Na osnovu informacija iz kolone InvoiceDate lako se određuje najskoriji datum, pa se tom datumu „dodaje“ još jedan dan i to uzima kao referentni datum. *R* – vrednost se zatim računa za svaku vrstu tj. za svaku transakciju iz skupa podataka kao razlika referentnog datuma i datuma kada je transakcija obavljena. *F* – vrednost se računa kao broj faktura po kupcu, predstavlja ukupan broj kupovina koje je kupac obavio. *M* – vrednost se određuje kao suma potrošene količine novca pri u svim kupovinama. Količina novca koju je kupac potrošio u jednoj transakciji se računa kao proizvod jedinične cene artikla i količine. Na slici 8 prikazane su neke od izračunatih vrednosti parametara *R*, *F* i *M*. Na slici 9 je prikazan dobijeni *RFM* model.

0	326	CustomerID	Frequency	CustomerID	TotalPrice	
1	2	0	12346.0	0	12346.0	77183.60
2	75	1	12347.0	1	12347.0	4310.00
3	19	2	12348.0	2	12348.0	1797.24
4	310	3	12349.0	3	12349.0	1757.55
		4	12350.0	4	12350.0	334.40

**Slika 8.** Recency, Frequency i Monetary vrednosti

	Unnamed: 0	CustomerID	Recency	Frequency	Monetary
count	4338.000000	4338.000000	4338.000000	4338.000000	4338.000000
mean	2168.500000	15300.408022	92.536422	4.272015	2054.266460
std	1252.417063	1721.808492	100.014169	7.697998	8989.230441
min	0.000000	12346.000000	1.000000	1.000000	3.750000
25%	1084.250000	13813.250000	18.000000	1.000000	307.415000
50%	2168.500000	15299.500000	51.000000	2.000000	674.485000
75%	3252.750000	16778.750000	142.000000	5.000000	1661.740000
max	4337.000000	18287.000000	374.000000	209.000000	280206.020000

Slika 9. *RFM* model

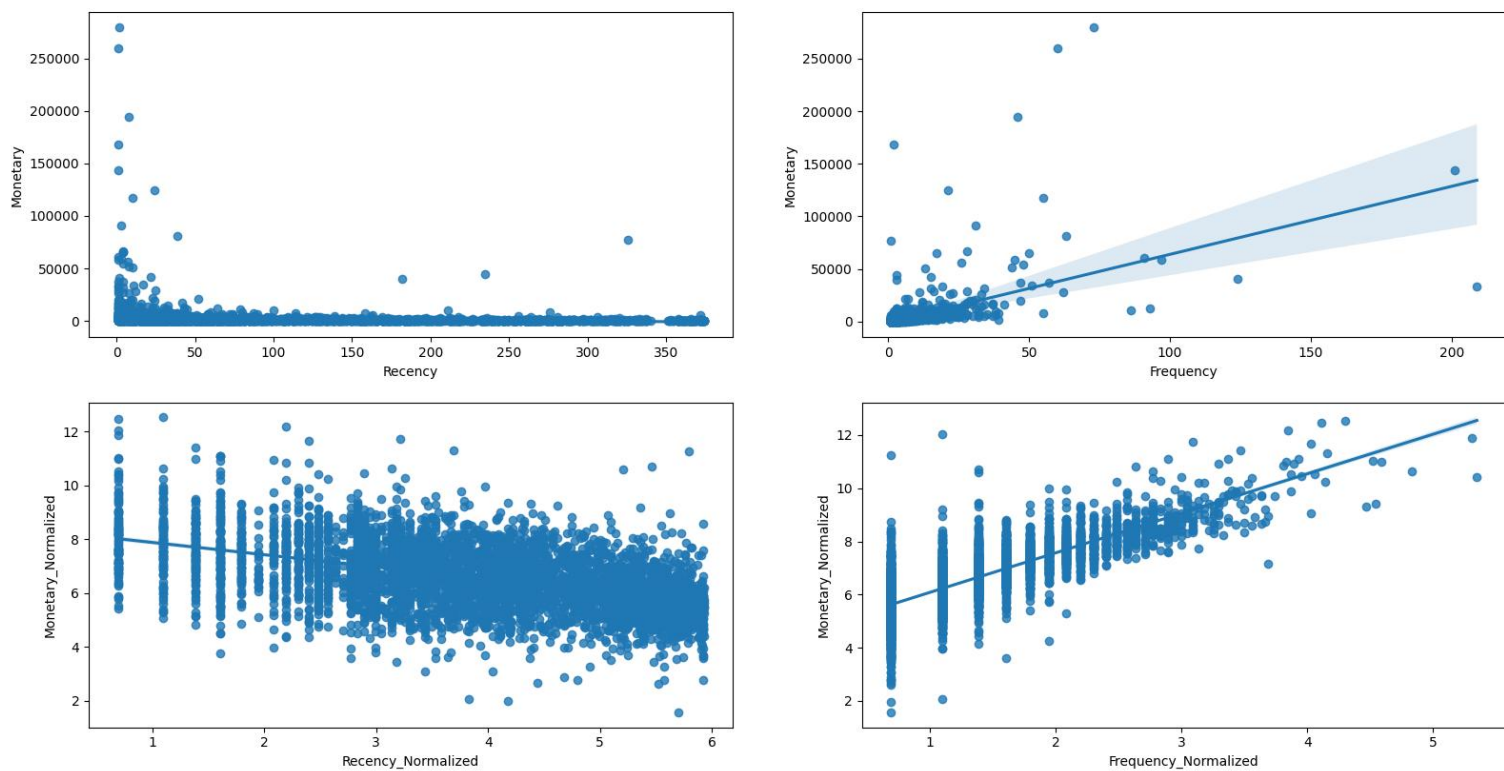
Segmentacija kupaca je izvršena primenom *K-Means* algoritma nad *RFM* modelom. Pre procesa klasterizacije, izvršena je logaritamska transformacija vrednosti u kolonama *Recency*, *Frequency* i *Monetary* primenom formule  $\log(1 + x)$ . Ova transformacija smanjuje asimetriju podataka, čime se približava normalnoj distribuciji – velike vrednosti se kompresuju, dok se manje vrednosti šire. Takva normalizacija podataka je važna jer asimetrični podaci mogu negativno uticati na performanse algoritame klasterizacije. Pored toga, vrednosti su standardizovane uklanjanjem srednje vrednosti i skaliranjem na jedinicu varijanse (srednja vrednost 0, standardna devijacija 1). Standardizacija je ključna, jer bez nje, karakteristike sa većim opsegom vrednosti preuzimaju primat u procesu klasterovanja, što dovodi do nepravilne segmentacije. Skalirane *RFM* vrednosti su prikazane na slici 10.

```
[ [ 1.46199281 -0.95521426 3.70622476 ]
  [-2.03873442 1.07442519 1.41184341 ]
  [ 0.37310424 0.38630445 0.7164889 ]
  ...
  [-1.21893976 -0.36158278 -1.11812113]
  [-1.65755161 2.17800394 0.83829669]
  [-0.03473174 0.05960547 0.73400231]]
```

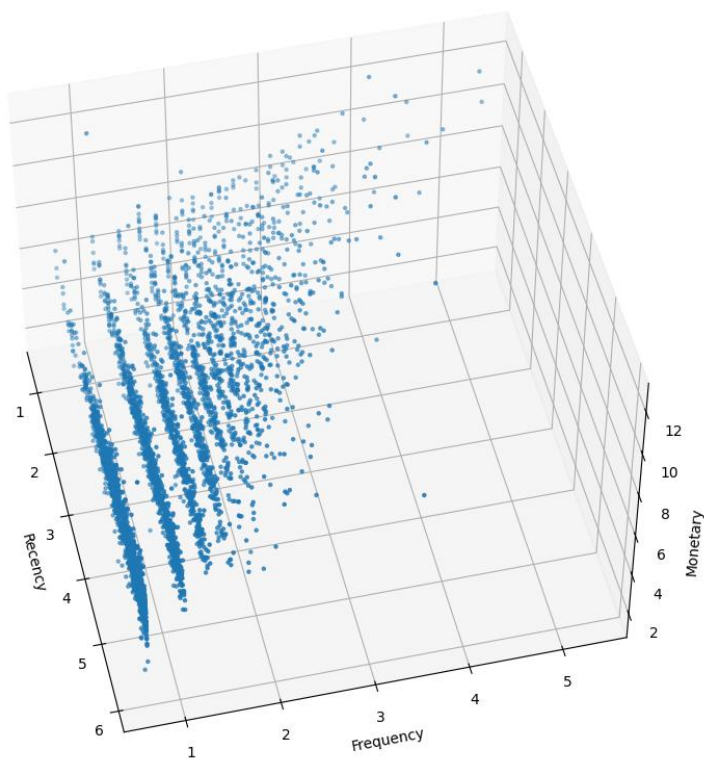
Slika 10. Skalirane *RFM* vrednosti

Na slici 11 prikazani su dijagrami raspodele R, F i M vrednosti. Prvobitno se primećuje negativna korelacija između R i M vrednosti – veće recency vrednosti su povezane sa manjim monetary vrednostima, što ukazuje na to da kupci koji su nedavno kupovali (niže recency vrednosti) ostvaruju veći prihod. S druge strane, postoji pozitivna korelacija između F i M vrednosti – kupci koji češće kupuju (više frequency vrednosti) u proseku troše više novca, što je očekivano. Nakon normalizacije ovih vrednosti, distribucija postaje uravnoteženija. Iako je negativna korelacija između R i M vrednosti i dalje prisutna, ona je manje izražena, dok pozitivan odnos između F i M vrednosti ostaje stabilan.





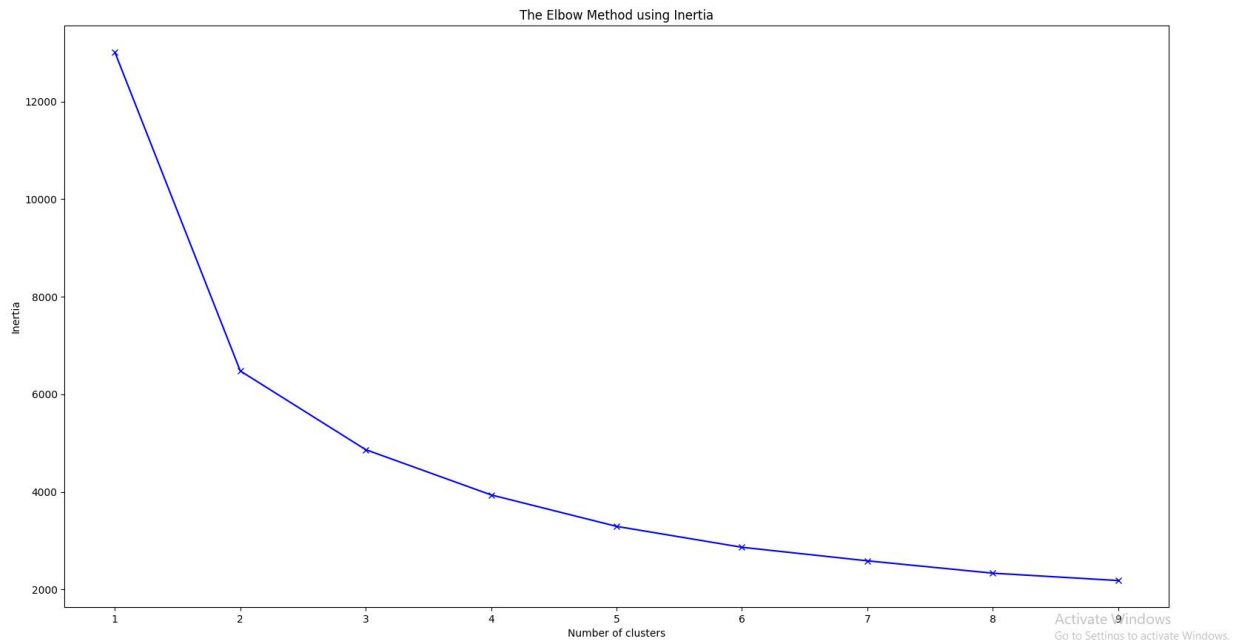
Slika 11. Distribucija *RFM* vrednosti



Slika 12. 3D distribucija *RFM* vrednosti



Za pronalaženje optimalnog broja klaster iskorišćen je *elbow* metod (slika 13). Uočava se da je optimalan broj klastera 3. Međutim, za potrebe segmentacije uzeto je da je broj klastera 4 budući da granice između prepoznatih kategorija kupaca nisu toliko stroge, te je samim tim korisno razlikovati više kategorija.



**Slika 13.** Određivanje optimalnog broja klastera

	CustomerID	Recency	Frequency	Monetary	Cluster
0	12346.0	326	1	77183.60	2
1	12347.0	2	7	4310.00	1
2	12348.0	75	4	1797.24	2
3	12349.0	19	1	1757.55	0
4	12350.0	310	1	334.40	3

**Slika 14.** Rezultati klasterizacije

Klasteri su inicijalno označeni vrednostima 0-3, pa je potrebno prepoznati koje su to kategorije kupaca prema podeli datoj u tabeli 5. Sa tim ciljem određuju se centriodi dobijenih klastera.

	Cluster	NumCustomers	Recency	Frequency	Monetary
0	0	837	18.124253	2.148148	551.819534
1	1	716	12.131285	13.713687	8074.266872
2	2	1173	71.084399	4.083546	1802.829005
3	3	1612	182.496898	1.318238	343.450032

**Slika 15.** Centroidi dobijenih klastera

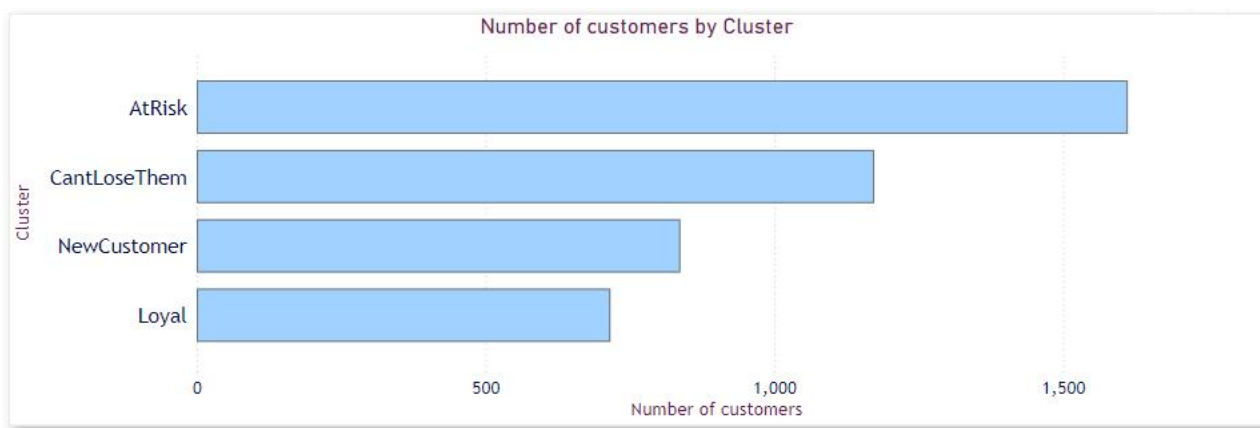
**Klaster 0.** Potrošači koji su nedavno izvršili kupovinu (u poslednjih dvadesetak), potrošili su pristojnu sumu novca, ali do sada nisu često kupovali. Ovi potrošači pripadaju kategoriji **NewCustomers**.

**Klaster 1.** Potrošači koji često kupuju, a poslednju kupovinu su obavili pre oko desetak dana, pri čemu troše velike svote novca. Ovi potrošači pripadaju kategoriji **Loyal**.

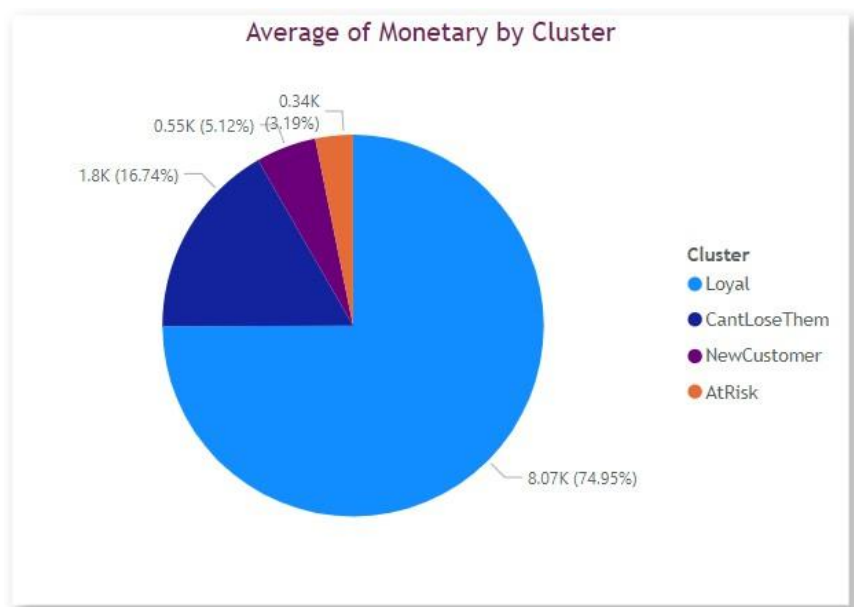
**Klaster 2.** Potrošači koji povremeno kupuju i nisu bili aktivni u poslednje vreme (poslednju kupovinu su obavili pre otprilike dva meseca), ali troše značajne svote novca. Ovi potrošači pripadaju kategoriji **Can't lose them but losing**.

**Klaster 3.** Potrošači koji nisu obavili kupovinu u poslednjih šest meseci, retko kupuju i troše male svote novca. Ovi potrošači pripadaju kategoriji **At risk**.

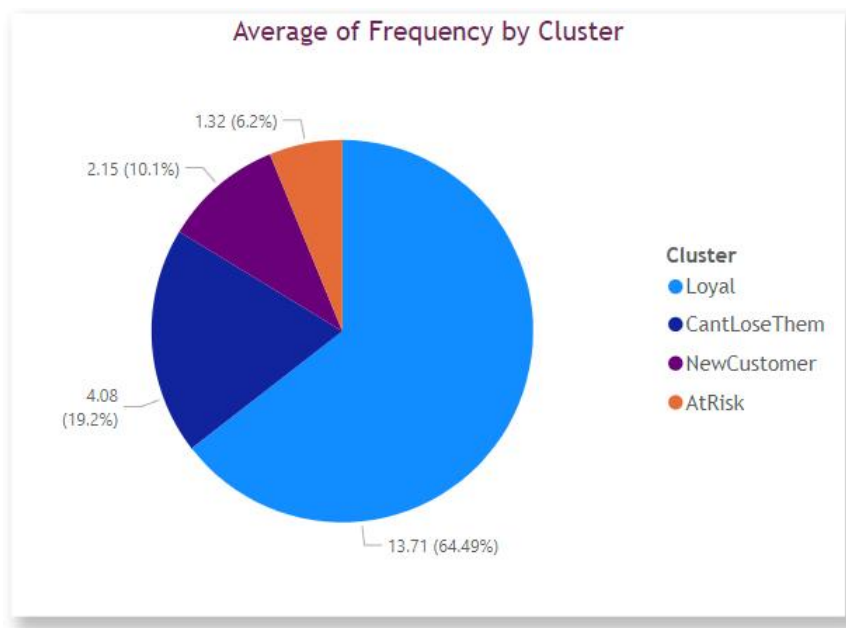
Za vizuelizaciju dobijenih rezultata korišćen je Microsoft-ov alat PowerBI. Podaci su uvezeni u *csv* (eng. *Comma Separated Value*) format. Sledeći grafikoni prikazuju klaster u odnosu na broj potrošača u njima, kao i u odnosu na svaku od vrednosti iz RFM modela i dokazuju prethodno donete zaključke.



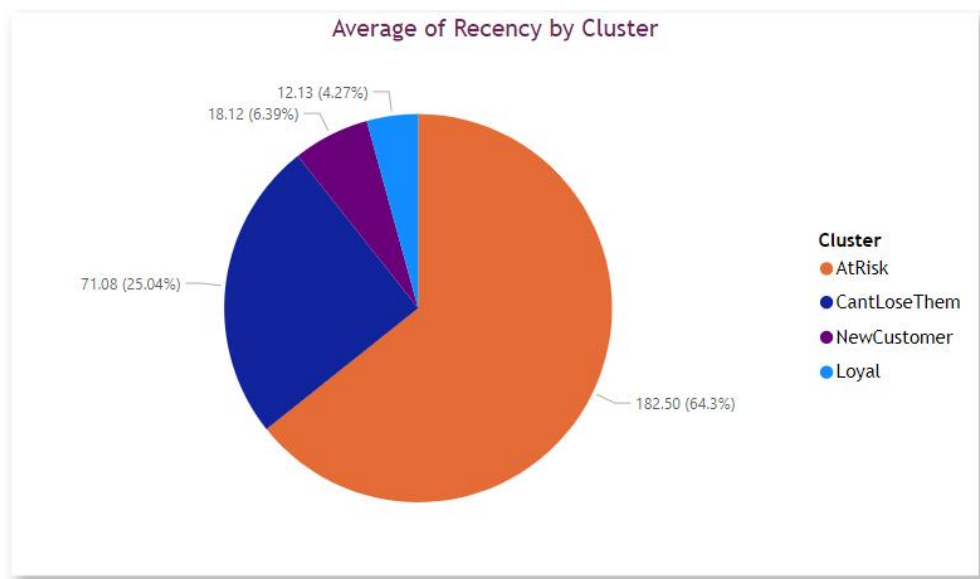
**Slika 16.** Broj potrošača po klasteru



**Slika 17.** Udeo klastera u ukupnoj monetarnoj vrednosti

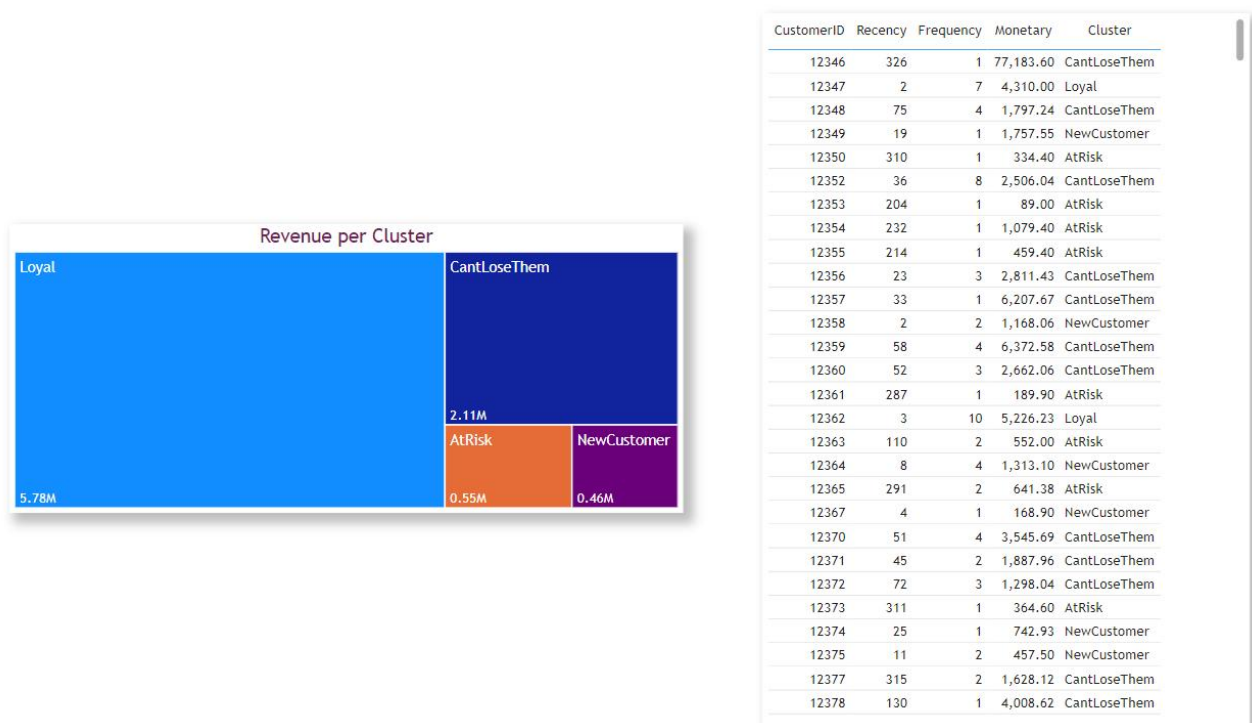


**Slika 18.** Učestanost kupovine po klasterima



**Slika 19.** Skorašnjost kupovine po klasterima

Na slici 20 su prikazani klasteri u formi grida. Veličina segmenata u gridu (koji predstavljaju klasteru) je određena na osnovu sume monetarnih vrednosti po klasteru. Ovaj kriterijum je odabran zato što je upravo monetarna vrednost tj. prihod koji kupci iz određenih kategorija donose kompaniji, najvažniji parametar i sve strategije za poboljšanje poslovanja koje se primenjuju kao krajnji cilj imaju povećanje monetarne vrednosti tj. dobiti od kupaca.



Slika 20. Rezultati segmentacije klijenata/kupaca

Za potrebe izračunavanja CLV vrednosti korišćen je skup podataka koji sadrži sve potrebne informacije za analizu prodaje i ponašanja kupaca, a preuzet je sa adrese: <https://www.kaggle.com/datasets/serdarsozturk/customer-segmentationrfmclvchurn>. Ovaj skup podataka sadrži sledeće informacije:

- invoiceID: jedinstveni identifikator fakture. Karakter 'C' na početku nekih vrednosti (npr C570727) ukazuje na fakture koje su stornirane ili vraćene.
- invoice\_date: datum i vreme izdavanja fakture, što može pomoći u analizi sezonskih trendova i ponašanja kupaca u različitim vremenskim periodima.
- customerID: jedinstveni ID kupca
- country: zemlja iz koje dolazi kupac, što može pomoći u analizi geografskih tržišta.
- quantity: količina proizvoda kupljena u svakoj transakciji. Negativna količina (npr -1) ukazuje na povrat ili storniranu transakciju.
- amount: jedinična cena u valuti (verovatno u funtama)

Slika 21a ilustruje primer neprečišćenih podataka iz korišćenog skupa podataka, dok su na slici 21b prikazani podaci nakon preprociranja.

	invoiceID	invoice_date	customerID	country	quantity	amount
0	548370	3/30/2021 16:14	15528.0000	United Kingdom	123	229,33
1	575767	11/11/2021 11:11	17348.0000	United Kingdom	163	209,73
2	0570727	10/12/2021 11:32	12471.0000	Germany	-1	-1,45
3	549106	4/6/2021 12:08	17045.0000	United Kingdom	1	39,95
4	573112	10/27/2021 15:33	16416.0000	United Kingdom	357	344,83
5	576630	11/16/2021 8:38	13816.0000	Germany	91	199,98

Slika 21a. – Neprečišćeni podaci iz tobacco\_sales dataset-a

	invoiceID	invoice_date	customerID	country	quantity	amount
0	548370	3/30/2021 16:14	15528.0000	United Kingdom	123.0000	229.3300
1	575767	11/11/2021 11:11	17348.0000	United Kingdom	163.0000	209.7300
3	549106	4/6/2021 12:08	17045.0000	United Kingdom	1.0000	39.9500
4	573112	10/27/2021 15:33	16416.0000	United Kingdom	357.0000	344.8300
5	576630	11/16/2021 8:38	13816.0000	Germany	91.0000	199.9800

Slika 21b. – Prečišćeni podaci iz tobacco\_sales dataset-a

Na početku se izračunava ukupna zarada po kupcu množenjem količine i cene po jedinici artikla. Zatim se računaju sledeći parametri (slika 22):

- **total\_transactions\_num**: ukupan broj jedinstvenih transakcija po kupcu.
- **total\_quantity**: ukupna količina proizvoda koju je svaki kupac kupio.
- **total\_income**: ukupan prihod ostvaren od svakog kupca.

customerID	total_transactions_num	total_quantity	total_income
12346.0000	1	5673.5000	54456908.1980
12347.0000	7	2458.0000	1839237.3600
12348.0000	4	2341.0000	1422255.6400
12349.0000	1	631.0000	1109014.0500
12350.0000	1	197.0000	65876.8000

Slika 22. Priprema parametara za CLV analizu

Na osnovu ovih podataka se izračunavaju dodatne metrike (slika 23):

- **average\_order\_value**: prosečna vrednost porudžbine, kao ukupan prihod podeljen brojem transakcija.
- **purchase\_frequency**: učestalost kupovine, kao odnos broja transakcija i ukupnog broja kupaca.
- **repeat\_rate**: procenat kupaca koji su napravili više od jedne transakcije.
- **churn\_rate**: procenat kupaca koji su prestali da kupuju.
- **profit\_margin**: profitna marža, kao trećina ukupnog Prihoda.
- **customer\_value**: Vrednost kupca, kao proizvod prosečne vrednosti porudžbine i učestalosti kupovine.
- **customer\_lifetime\_value**: Procena vrednosti kupca tokom vremena, kao vrednost kupca podeljena churn rate-om, a zatim pomnožena sa profitnom maržom.

	total_transactions_num	total_quantity	total_income	average_order_value	purchase_frequency	profit_margin	customer_value	customer_lifetime_value	class
customerID									
14646.0000	73	152084.5000	1058537852.5900	14500518.5286	0.0168	352845950.8633	244015.1804	250167980657702.6250	A
18102.0000	60	64124.0000	420707241.9800	7011787.3663	0.0138	140235747.3267	96981.8446	39516540177365.1562	A
12415.0000	21	50647.0000	393877468.5900	18756069.9332	0.0048	131292489.5327	90797.0190	34637075300104.6758	A
17450.0000	46	62008.0000	390982171.4940	8499612.4238	0.0106	130327390.4980	90129.5923	34129729499031.9141	A
14156.0000	55	57308.0000	288281262.2340	5241477.4952	0.0127	96093754.0780	66454.8783	18554607313067.3047	A

Slika 23. Lista kupaca u opadajućem poretku prema izračunatoj CLV vrednosti

class	total_transactions_num			total_quantity			total_income		
	count	sum	mean	count	sum	mean	count	sum	mean
D	1085	1663	1.5327	1085	106741.0000	98.3788	1085	16560378.4900	15263.0216
C	1084	2471	2.2795	1084	296479.0000	273.5046	1084	79405036.3200	73251.8785
B	1084	4309	3.9751	1084	724910.0000	668.7362	1084	273820372.7000	252601.8198
A	1085	10089	9.2986	1085	3795613.0000	3498.2608	1085	7463088629.3890	6878422.6999
average_order_value	purchase_frequency			profit_margin					
	count	sum	mean	count	sum	mean	count	sum	mean
	1085	12743634.6954	11745.2854	1085	0.3834	0.0004	1085	5520126.1633	5087.6739
	1084	49562403.1676	45721.7741	1084	0.5696	0.0005	1084	26468345.4400	24417.2928
	1084	114474316.5559	105603.6131	1084	0.9933	0.0009	1084	91273457.5667	84200.6066
	1085	842151499.9857	776176.4977	1085	2.3257	0.0021	1085	2487696209.7963	2292807.5666
customer_value	customer_lifetime_value								
	count	sum	mean	count	sum	mean	count	sum	mean
	1085	3817.5146	3.5184	1085	80874008.7810	74538.2569			
	1084	18304.5266	16.8861	1084	1461318702.0655	1348079.9835			
	1084	63121.3399	58.2300	1084	17552007081.2786	16191888.4514			
	1085	1720398.4853	1585.6207	1085	450877873075438.1875	415555643387.5006			

Slika 24. Statistika po klasterima

Na osnovu prikazane statistike može se zaključiti da su vrednosti koje pripadaju klasi ‘A’ višestruko veće od vrednosti u ostalim klasterima. To bi značilo da su verovatno kupci koji pripadaju ovoj kategoriji najznačajniji za kompaniju i da joj donose najveću vrednost.

**Priprema podataka za BG-NBD model:** Podaci se filtriraju tako da obuhvataju samo kupce sa više od jedne transakcije. Zatim se izračunavaju sledeće varijable za BG-NBD model:

- **R:** vreme između prve i poslednje transakcije kupca (u danima, podeljeno na nedelje).
- **T:** starost kupca (vreme od prve transakcije do trenutnog datuma, u nedeljama).
- **F:** broj jedinstvenih transakcija po kupcu.
- **M:** prosečna zarada po transakciji, kao ukupan prihod podeljen brojem transakcija.

Zatim se vrši treniranje **BG-NBD** modela pomoću prethodno izračunatih parametara za predviđanje broja transakcija po kupcu u narednom periodu i **Gamma-Gamma** modela za predviđanje prosečne zarade po transakciji po kupcu. Na posletku, ova dva modela se kombinuju za izračunavanje konačne CLV vrednosti kupaca. Rezultati su prikazani na slici 25.



	R	T	F	M	expected_purchase_1_month	expected_purchase_3_months	expected_average_profit	clv
customerID								
14646.0000	50.4286	50.8571	73	14500518.5286	4.7964	14.2869	14573438.9990	221533751.0428
12415.0000	44.7143	48.4286	21	18756069.9332	1.5074	4.4883	19088081.5001	91151855.7997
18102.0000	52.2857	52.5714	60	7011787.3663	3.8617	11.5046	7054735.0691	86357412.0697
17450.0000	51.2857	52.7143	46	8499612.4238	2.9711	8.8513	8567644.5458	80689489.2523
16000.0000	0.0000	0.5714	3	12595687.6667	1.6371	4.7656	14341893.8702	72558601.7822
14156.0000	51.5714	53.2857	55	5241477.4952	3.4868	10.3884	5276520.0301	58323998.6933
14096.0000	13.8571	14.7143	17	5842932.4688	2.8729	8.4848	5971232.1528	53858801.4619
16684.0000	50.4286	51.2857	28	8270320.8206	1.9059	5.6768	8379635.2521	50613411.0312
17511.0000	52.8571	53.5714	31	7682823.6871	2.0236	6.0289	7774428.2717	49871569.4652
16446.0000	29.1429	29.5714	2	27228456.9990	0.3915	1.1612	33312416.3815	41138058.1279

**Slika 25.** Rezultati predikcije broja transakcija po kupcu u narednih 1 i 3 meseca (BG/NDB), i prosečnog prihoda (Gamma)

- **expected\_purchase\_1\_month** – predstavlja predikciju broja kupovina koje će svaki kupac obaviti u narednih mesec dana. Na primer, vrednost 4.7964 označava procenu da će kupac obaviti više od 4 kupovine za mesec dana, što ga čini vrednim za kompaniju, budući da je aktivan i da će verovatno nastaviti da kupuje. Nasuprot tome, neki kupci imaju niska predviđanja (0.3915) što sugerise da su ti kupci manje aktivni i da će verovatno prestati da kupuju i donose profit kompaniji.
- **expected\_purchase\_3\_months** – predstavlja predikciju broja kupovina koje će svaki kupac obaviti u naredna 3 meseca. Očekivano, vrednosti su veće jer obuhvataju duži period. Kupci sa visokim vrednostima u ovoj koloni (npr. 14.29 ili 11.50) su verovatno angažovaniji i imaju tendenciju da često kupuju. Kupci sa nižim brojevima (npr. 1.16) će kupovati povremeno, ali nisu ključni u smislu stalnog prometa.
- **expected\_average\_profit** – pruža uvid u prosečnu zaradu koju kupci donose kompaniji pri svakoj kupovini. Ova metrika može pomoći da se proceni koliko su određeni kupci profitabilni za kompaniju u budućnosti. Veće vrednosti označavaju veći profit tj. zaradu koju kompanija ostvaruje, te su takvi kupci značajniji i treba ih zadržati.
- **CLV** – predstavlja ukupnu vrednost koju kupac može doneti kompaniji tokom celog svog angažmana. Ova vrednost kombinuje očekivani broj budućih kupovina i prosečan profit po kupovini. Kupci sa visokim CLV (npr. 221 533 751) su ključni za poslovanje jer donose veliku vrednost na duži period. Oni su idealni kandidati za zadržavanje i lojalnost, jer će kompaniji verovatno donositi najviše prihoda u budućnosti. Kupci sa nižim CLV vrednostima (npr. 411 380 58) nisu profitabilni na duže staze, što može značiti da nisu prioritet za strategije zadržavanja ili mogu biti u fokusu marketinških kampanja za ponovno angažovanje.

Drugi deo praktične implementacije se bavi analizom potrošačke korpe. U ovu svrhu je korišćen ranije pomenuti online-retail skup podataka. Za formiranje asocijativnih pravila korišćen je Apriori algoritam. Na osnovu podataka o stavkama sadržanim u transakcijama (kolona Description) formira se binarna matrica u kojoj svaka vrsta predstavlja pojedinačnu transakciju kolone predstavljaju stavke koje su sadržane u svim transakcijama (prethodno se formira jedinstveni skup stavki iz svih transakcija). U preseku odabrane vrste i kolone vrednost će biti 1 ukoliko se posmatrana stavka nalazi u datoj transakciji, u suprotnom će vrednost biti 0.

Za potrebe analize potrošačke korpe korišćen je online-retail skup podataka. Za formiranje asocijativnih pravila korišćen je *Apriori* algoritam. Na osnovu podataka o stavkama sadržanim u transakcijama (kolona *Description*) formira se binarna matrica u kojoj svaka vrsta predstavlja pojedinačnu transakciju kolone predstavljaju stavke koje su sadržane u svim transakcijama (prethodno se formira jedinstveni skup stavki iz svih transakcija). U preseku odabrane vrste i kolone vrednost će biti 1 ukoliko se posmatrana stavka nalazi u datoj transakciji, u suprotnom će vrednost biti 0.

	WHITE HA	WHITE ME	CREAM CL	KNITTED U	RED WOO	SET 7 BAB	GLASS STA	HAND WA	HAND WA	ASSORTE	POPPY'S P	POPPY'S P	FELTCRAF
0	1	1	1	1	1	1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	1	0	0	0	0
2	0	0	0	0	0	0	0	0	0	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0

Slika 26. Matrica transakcija

Naredni korak je formiranje frekventnog skupa podataka. Vrednost *min\_support* parametra je postavljena na 0,01 što znači da stavka mora da učestvuje u najmanje 1% transakcija kako bi bila uzeta u razmatranje. Dobijeni frekventni skupovi (slika 27) se koriste za formiranje asocijativnih pravila.

	support	itemsets
0	0.106357	(WHITE HANGING HEART T-LIGHT HOLDER)
1	0.012843	(WHITE METAL LANTERN)
2	0.013005	(CREAM CUPID HEARTS COAT HANGER)
3	0.017213	(KNITTED UNION FLAG HOT WATER BOTTLE)
4	0.017321	(RED WOOLLY HOTTIE WHITE HEART.)
..	...	...
968	0.014300	(ROSES REGENCY TEACUP AND SAUCER , REGENCY CAK...
969	0.014623	(GREEN REGENCY TEACUP AND SAUCER, REGENCY CAKE...
970	0.021045	(ROSES REGENCY TEACUP AND SAUCER , GREEN REGEN...
971	0.010414	(LUNCH BAG RED RETROSPOT, LUNCH BAG CARS BLUE,...
972	0.012897	(ROSES REGENCY TEACUP AND SAUCER , GREEN REGEN...

Slika 27a. Frekventni skupovi stavki – *Apriori*

	support	itemsets
0	0.106357	(WHITE HANGING HEART T-LIGHT HOLDER)
1	0.017321	(RED WOOLLY HOTTIE WHITE HEART.)
2	0.017213	(KNITTED UNION FLAG HOT WATER BOTTLE)
3	0.016026	(SET 7 BABUSHKA NESTING BOXES)
4	0.013005	(CREAM CUPID HEARTS COAT HANGER)
..	...	...
968	0.010792	(LUNCH BAG BLACK SKULL., LUNCH BAG VINTAGE DO...
969	0.010253	(LUNCH BAG SPACEBOY DESIGN , LUNCH BAG VINTAGE...
970	0.010199	(LUNCH BAG SUKI DESIGN , LUNCH BAG VINTAGE DOI...
971	0.011763	(HOT WATER BOTTLE KEEP CALM, LOVE HOT WATER BO...
972	0.010954	(HAND WARMER RED LOVE HEART, HAND WARMER OWL D...

Slika 27b. Frekventni skupovi stavki – *FPGrowth*



Evaluacija pravila se vrši primenom lift metrike. Ova metrika se koristi kako bi se procenila “jačina veze” između artikala. Zapravo, lift metrika procenjuje koliko je verovatno da se dve stavke pojave zajedno u transakciji u poređenju sa slučajem kada bi bile međusobno nezavisne. Parametar *min\_threshold* ima vrednost 1, čime se uzimaju u razmatranje pravila koja označavaju veću verovatnoću da se stavke iz pravila da se nađu zajedno u transakciji.

Iz skupa pravila izdvojena su ona koja zadovoljavaju uslov minimalne pouzdanosti (u kodu 0.7) i čija je vrednost za lift veća od 5, kako bi se izdvojila samo najznačajnija pravila. Na slici 28 je prikazano nekoliko pravila koji ispunjavaju ove uslove. Vreme potrebno za formiranje asocijativnih pravila je prilično brže kada se koristi *FPGrowth* algoritam što potvrđuje preporuku za njegovo korišćenje nad velikim skupovima podataka kao što je korišćeni.

id	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
14	POPPY'S PLAYHOUSE KITCHEN	POPPY'S PLAYHOUSE BEDROOM	0.018670408	0.017051586	0.013706022	0.734104046	43.05195
15	POPPY'S PLAYHOUSE BEDROOM	POPPY'S PLAYHOUSE KITCHEN	0.017051586	0.018670408	0.013706022	0.803797468	43.05195
27	ALARM CLOCK BAKELIKE PINK	ALARM CLOCK BAKELIKE RED	0.018508526	0.047323548	0.014407511	0.778425656	16.44901
80	JUMBO STORAGE BAG SUKI	JUMBO BAG RED RETROSPOT	0.015162961	0.086337147	0.011871358	0.782918149	9.068149
86	JUMBO STORAGE BAG SUKI	JUMBO BAG RED RETROSPOT	0.013652061	0.086337147	0.010090654	0.739130435	8.560978
128	SET/6 RED SPOTTY PAPER PLATES	SET/20 RED RETROSPOT PAPER NAPKINS	0.017537233	0.039499245	0.012357004	0.704615385	17.83871
144	JUMBO BAG PINK POLKADOT	JUMBO BAG RED RETROSPOT	0.013166415	0.086337147	0.010144615	0.770491803	8.924221
184	LUNCH BAG DOLLY GIRL DESIGN	LUNCH BAG SPACEBOY DESIGN	0.013652061	0.053313188	0.010090654	0.739130435	13.86393
202	SET/6 RED SPOTTY PAPER CUPS	SET/6 RED SPOTTY PAPER PLATES	0.015378804	0.017537233	0.012734729	0.828070175	47.21784
203	SET/6 RED SPOTTY PAPER PLATES	SET/6 RED SPOTTY PAPER CUPS	0.017537233	0.015378804	0.012734729	0.726153846	47.21784
227	ALARM CLOCK BAKELIKE IVORY	ALARM CLOCK BAKELIKE RED	0.014839197	0.047323548	0.011601554	0.781818182	16.5207
308	LUNCH BAG BLACK SKULL	LUNCH BAG PINK POLKADOT	0.014137708	0.050237427	0.010414418	0.736641221	14.6632
310	LUNCH BAG BLACK SKULL	LUNCH BAG RED RETROSPOT	0.014029786	0.069501403	0.010414418	0.742307692	10.68047

Slika 28a. Generisana asocijativna pravila – *Apriori*

id	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
20	CANDLEHOLDER PINK HANGING HEART	WHITE HANGING HEART T-LIGHT HOLDER	0.018508526	0.106356572	0.013652061	0.737609329	6.935249
40	PAINTED METAL PEARS ASSORTED	ASSORTED COLOUR BIRD ORNAMENT	0.018994172	0.074195985	0.013759983	0.724431818	9.76376
44	POPPY'S PLAYHOUSE KITCHEN	POPPY'S PLAYHOUSE BEDROOM	0.018670408	0.017051586	0.013706022	0.734104046	43.05195
45	POPPY'S PLAYHOUSE BEDROOM	POPPY'S PLAYHOUSE KITCHEN	0.017051586	0.018670408	0.013706022	0.803797468	43.05195
46	POPPY'S PLAYHOUSE LIVINGROOM	POPPY'S PLAYHOUSE BEDROOM	0.013598101	0.017051586	0.011061947	0.813492063	47.70771
48	POPPY'S PLAYHOUSE LIVINGROOM	POPPY'S PLAYHOUSE KITCHEN	0.013598101	0.018670408	0.011601554	0.853174603	45.69662
75	ALARM CLOCK BAKELIKE CHOCOLATE	ALARM CLOCK BAKELIKE RED	0.017375351	0.047323548	0.012195122	0.701863354	14.83116
157	SET/6 RED SPOTTY PAPER PLATES	SET/20 RED RETROSPOT PAPER NAPKINS	0.017537233	0.039499245	0.012357004	0.704615385	17.83871
158	SET/6 RED SPOTTY PAPER CUPS	SET/6 RED SPOTTY PAPER PLATES	0.015378804	0.017537233	0.012734729	0.828070175	47.21784
159	SET/6 RED SPOTTY PAPER PLATES	SET/6 RED SPOTTY PAPER CUPS	0.017537233	0.015378804	0.012734729	0.726153846	47.21784
361	TOILET METAL SIGN	BATHROOM METAL SIGN	0.016296136	0.020666954	0.012141161	0.745033113	36.04949
454	WOODEN TREE CHRISTMAS SCANDINAVIAN	WOODEN HEART CHRISTMAS SCANDINAVIAN	0.014893158	0.024983812	0.010792143	0.724637681	29.00429
456	WOODEN STAR CHRISTMAS SCANDINAVIAN	WOODEN HEART CHRISTMAS SCANDINAVIAN	0.02336499	0.024983812	0.017213469	0.736720554	29.48792

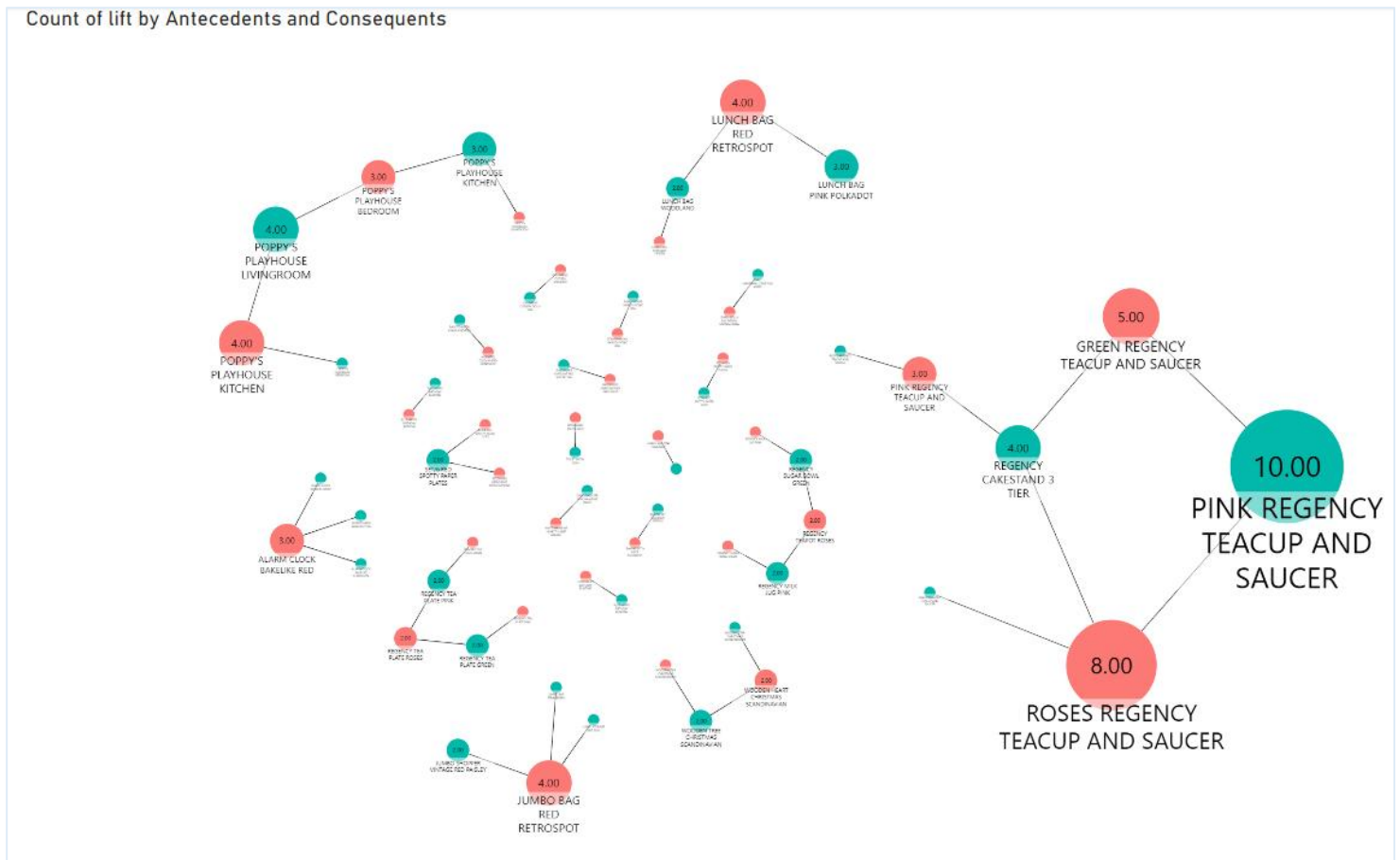
Slika 28b. Generisana asocijativna pravila – *FPGrowth*

Dobijena pravila su uneta su **PowerBI**, a onda je dat njihov tabelarni prikaz (slika 29). Kolona *Antecedents* označava levu stranu pravila dok kolona *Consequents* označava desnu stranu pravila. To znači da potrošači često, prilikom kupovine stavki iz kolone *Antecedents* kupuju i stavke iz kolone *Consequents*, te shodno tome se daje preporuka da se u prodavnicama ovi proizvodi postavljaju na susednim policama, kako bi kupcima bili uočljiviji.

Antecedents	Consequents	support	confidence	lift
ALARM CLOCK BAKELIKE CHOCOLATE	ALARM CLOCK BAKELIKE RED	0.01	0.70	14.83
ALARM CLOCK BAKELIKE IVORY	ALARM CLOCK BAKELIKE RED	0.01	0.78	16.52
ALARM CLOCK BAKELIKE PINK	ALARM CLOCK BAKELIKE RED	0.01	0.78	16.45
BAKING SET SPACEBOY DESIGN	BAKING SET 9 PIECE RETROSPOT	0.02	0.73	15.92
BLUE HAPPY BIRTHDAY BUNTING	PINK HAPPY BIRTHDAY BUNTING	0.01	0.72	36.93
CANDLEHOLDER PINK HANGING HEART	WHITE HANGING HEART T-LIGHT HOLDER	0.01	0.74	6.94
CHILDRENS CUTLERY DOLLY GIRL	CHILDRENS CUTLERY SPACEBOY	0.01	0.77	45.30
GARDENERS KNEELING PAD CUP OF TEA	GARDENERS KNEELING PAD KEEP CALM	0.02	0.73	17.87
GREEN REGENCY TEACUP AND SAUCER	ROSES REGENCY TEACUP AND SAUCER	0.03	0.78	18.53
JUMBO BAG STRAWBERRY	JUMBO BAG RED RETROSPOT	0.01	0.79	9.17
JUMBO SHOPPER VINTAGE RED PAISLEY	JUMBO BAG RED RETROSPOT	0.01	0.74	8.56
JUMBO SHOPPER VINTAGE RED PAISLEY	JUMBO BAG RED RETROSPOT	0.01	0.77	8.92
JUMBO STORAGE BAG SUKI	JUMBO BAG RED RETROSPOT	0.01	0.78	9.07
LUNCH BAG CARS BLUE	LUNCH BAG PINK POLKADOT	0.01	0.74	14.66
LUNCH BAG PINK POLKADOT	LUNCH BAG RED RETROSPOT	0.01	0.74	10.68
LUNCH BAG PINK POLKADOT	LUNCH BAG RED RETROSPOT	0.01	0.75	10.72
LUNCH BAG PINK POLKADOT	LUNCH BAG RED RETROSPOT	0.01	0.70	10.13
LUNCH BAG WOODLAND	LUNCH BAG RED RETROSPOT	0.01	0.72	10.40
LUNCH BAG WOODLAND	LUNCH BAG SPACEBOY DESIGN	0.01	0.74	13.86
PAINTED METAL PEARS ASSORTED	ASSORTED COLOUR BIRD ORNAMENT	0.01	0.72	9.76
PINK HAPPY BIRTHDAY BUNTING	BLUE HAPPY BIRTHDAY BUNTING	0.01	0.71	36.93
PINK REGENCY TEACUP AND SAUCER	GREEN REGENCY TEACUP AND SAUCER	0.01	0.90	24.19
PINK REGENCY TEACUP AND SAUCER	GREEN REGENCY TEACUP AND SAUCER	0.01	0.88	23.52
PINK REGENCY TEACUP AND SAUCER	GREEN REGENCY TEACUP AND SAUCER	0.02	0.89	23.99
PINK REGENCY TEACUP AND SAUCER	GREEN REGENCY TEACUP AND SAUCER	0.02	0.83	22.19
PINK REGENCY TEACUP AND SAUCER	ROSES REGENCY TEACUP AND SAUCER	0.01	0.77	26.50

**Slika 29.** Tabelarni prikaz asocijativnih pravila

Na slici 30 prikazana su asocijativna pravila dobijena primenom *Apriori* algoritma u vidu mreže. Veličina čvora u mreži odgovara vrednosti *lift*-a za konkretno pravilo. Ilustrovana su samo najpouzdanija pravila tj. pravila kod kojih je vrednost *lift*-a veća od 0.7.



**Slika 30.** Mreža asocijativnih pravila

## 4. Zaključak

Promenljivo tržišno okruženje zahteva od kompanija da se kontinuirano prilagođavaju kako bi ostale konkurentne. Zato je od velike važnosti da kompanije pažljivo analiziraju ostvarene rezultate i, na osnovu tih podataka, definišu buduće poslovne strategije. U tom kontekstu, poslovna inteligencija igra ključnu ulogu, omogućavajući organizacijama da prikupljaju, istražuju, obrađuju i vizualizuju podatke, čime se olakšava donošenje informisanih odluka.

Analitika, kao proces razvoja preporuka i odluka zasnovanih na istorijskim podacima, predstavlja osnovu za strateško delovanje. Postoje tri osnovna tipa poslovne analitike: deskriptivna, prediktivna i preskriptivna. Deskriptivna analitika je prvi korak u analizi podataka, pružajući jasnu i razumljivu sliku prikupljenih informacija. Prediktivna analitika koristi statističke metode i tehnike otkrivanja znanja kako bi predvidela buduće trendove, dok preskriptivna analitika nastoji da pruži preporuke za optimalno donošenje odluka, uzimajući u obzir trenutne i buduće scenarije.

Jedan od ključnih aspekata deskriptivne analitike je segmentacija klijenata, koja omogućava identifikaciju profitabilnih i neprofitabilnih kupaca, te usmeravanje resursa i marketinških aktivnosti ka najvažnijim ciljnim grupama. RFM model, koji se fokusira na biheviorističke podatke, nudi uvid u kupovne navike i ponašanje potrošača. Pored toga, RFM model se često koristi kao osnova za procenu vrednosti kupca (CLV), što omogućava kompanijama da predvide buduće prihode od određenih kategorija kupaca. Osim toga, analiza potrošačke korpe pruža dragocene informacije o kombinacijama proizvoda koje se često kupuju zajedno, što omogućava kompanijama da razviju preciznije marketinške strategije i bolje zadovolje potrebe svojih kupaca. Korišćenjem ovih analiza, kompanije mogu ne samo da optimizuju svoje poslovne procese već i da unaprede korisničko iskustvo, što vodi dugoročnoj lojalnosti potrošača.

## 5. Literatura

- [1] W. Grossman, S. Rinderle-Ma, (2015), *Fundamentals of Business Intelligence*, Springer, (septembar 2024)
- [2] R. Sharda, D. Delen, E. Turban, (2018), *Business Intelligence, Analytics and Data Science*, Pearson, (septembar 2024)
- [3] R. Sherman, (2014), *Business Intelligence Guidebook – From Data Integration to Analytics*, Kaufmann, (septembar 2024)
- [4] C. Vercellis, (2009), *Business Intelligence – Data Mining and Optimization for Decision Making*, Wiley, (septembar 2024)
- [5] Tableau, (2024), *Business Intelligence – A complete overview*, [link]: <https://www.tableau.com/learn/articles/business-intelligence>, (septembar 2024)
- [6] M. Schniederjans, D. Schniederjans, C. Starkey, (2014), *Business Analytics – Principles, Concepts and Applications: What, Why and How*, Ft Pr, (septembar 2024)
- [7] UNSW Online, (2020), *Descriptive, Predictive & Prescriptive Analytics: What are the differences?*, [link]: <https://studyonline.unsw.edu.au/blog/descriptive-predictive-prescriptive-analytics>, (septembar 2024)
- [8] G. Blokdik, (2018), *Customer Segmentation*, 2<sup>nd</sup> edition, 5StarCooks, (septembar 2024)
- [9] Qualtrics, (2024), *What is customer segmentation analysis, and how it can help?*, [link]: <https://www.qualtrics.com/en-au/experience-management/brand/customer-segmentation/>, (septembar 2024)
- [10] Qualtrics, (2024), *Market Segmentation – definition, types, benefits and best practices*, [link]: <https://www.qualtrics.com/experience-management/brand/what-is-market-segmentation/>, (septembar 2024)
- [11] CleverTap, (2024), *RFM analysis for Customer Segmentation*, [link]: <https://clevertap.com/blog/rfm-analysis/>, (septembar 2024)
- [12] Loyoly, (2024), *How to use the RFM Segmentation to segment your customer base*, [link]: <https://www.loyoly.io/blog/rfm-segmentation>, (septembar 2024)
- [13] Qualtrics, (2024), *How to calculate Customer lifetime value?* [link]: <https://www.qualtrics.com/experience-management/customer/how-to-calculate-customer-lifetime-value/>, (septembar 2024)
- [14] Medium, (2024), *Customer Life Time Value Prediction by Using BG-NBD & Gamma-Gamma Models*, [link]: <https://medium.com/analytics-vidhya/customer-life-time-value-prediction-by-using-bg-nbd-gamma-gamma-models-and-applied-example-in-997a5ee481ad>



- [15] AnalyticsLabs, (2024), *Clustering*, [link]: <https://www.analytixlabs.co.in/blog/types-of-clustering-algorithms/>
- [16] GeeksForGeeks, (2024), *Market Basket Analysis in Data Mining*, [link]: <https://www.geeksforgeeks.org/market-basket-analysis-in-data-mining/>, (septembar 2024)
- [17] Microsoft, (2024), PowerBI, <https://powerbi.microsoft.com/en-us/>
- [18] A.Sonule, V.Menhe, (2023), *Customer segmentation using artificial neural network*, IJCRT, (septembar 2024)
- [19] A.Bhargav, R.P.Mathur, (2014), *Market basket analysis using artificial neural network*, IEEE, (septembar 2024)