



8. Osnove poslovne analitike uključujući R i SQL

Šta je poslovna analitika (engl. *business analytics* - BA)? Koje probleme rešava i koje alate koristi? Šta su R i SQL? Kako je BA povezan s R i SQL? Postoje li primeri dobre prakse u kojima se ti softveri koriste za rešavanje logističkih poslovnih problema?

Na ova i slična pitanja pokušaćemo dati odgovore u sledećem poglavlju.

8.1 Šta je poslovna analitika?

BA predstavlja holistički pristup analizi podataka i poslovnom odlučivanju. To je okruženje vođeno podacima s ciljem poboljšanja poslovnih performansi kompanije pružanjem temelja za informisanije donošenje odluka. To je sistemski proces razmišljanja koji primenjuje kvalitativne, kvantitativne i statističke računarske alate i metode za analizu podataka, sticanje uvida, informisanje i podršku donošenju odluka. Svaka određena analiza može da koristi različite tehnike uključujući dijagnostičke, prediktivne, preskriptivne i optimizacione modele (Power i dr., 2018). Mikalef i dr., (2019) daju plan za akademsko istraživanje i praktičnu primenu, ističući transformativni potencijal analitike kada se pravilno integriše u organizacione procese. U skladu s tim, autori navode da BA zahteva od organizacija da radikalno redizajniraju način na koji se takvim inicijativama pristupa, kako se dizajniraju i usavršavaju, kako se planiranje resursa i orkestriranje izvršava i strateški usklađuje, kao i da ponovno vrednuju svoje očekivane rezultate izvođenja, njihovu povezanost sa strateškim ciljevima i, kao rezultat toga, razviju odgovarajuće KPI-eve (Mikalef et al., 2019).

Glavni zadaci BA su osigurati kanal znanja kako bi se osigurala koherentna veza između sirovih podataka i poslovnih odluka. Opšti cilj je poslovna učinkovitost kroz 'vertikalizaciju', upotrebljivost i integraciju s operativnim sistemima (Kohavi i dr., 2002). BA ima mnogo područja primene i povezanih izvedenica: finansijska analitika, analitika lanca snabdevanja, analitika krize, analitika znanja, marketinška analitika, analitika kupaca, analitika usluga, analitika ljudskih resursa, analitika talenata, analitika procesa, analitika rizika (Holsapple i dr., 2014.) .



Postoje tri vrste platformi poslovne analitike:

- **Opisne** – posmatraju postojeće podatke i daju sažetak statistike i osnovnu vizualizaciju.
- **Prediktivne** – koriste postojeće podatke za procenu najverovatnijih budućih scenarija.
- **Preskriptivne** – automatski obrađuju veliku količinu podataka (engl. *big data*), poslovna pravila, tržišne uslove itd. Ove platforme koriste metode mašinskog učenja i veštačke inteligencije. Cilj je potpuno automatizovano donošenje odluka o tome koje akcije kompanija treba preduzeti s obzirom na trenutnu situaciju kako bi postigla željene poslovne ciljeve.

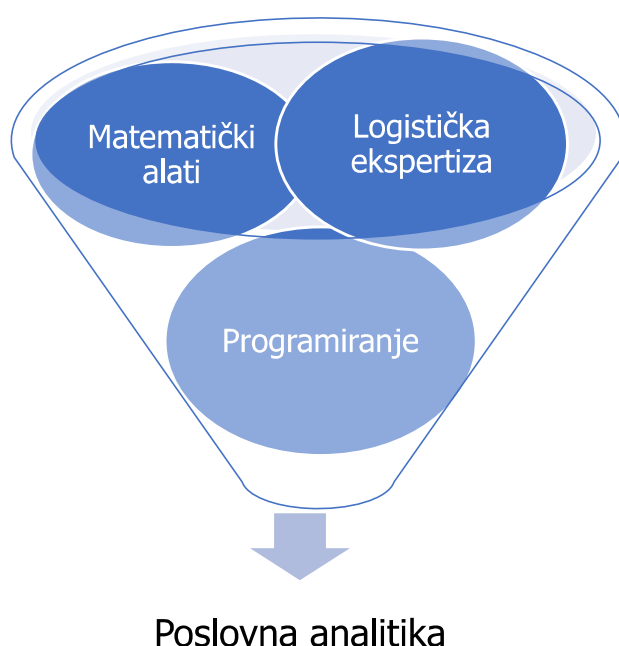
Zanimanje za *big data* i poslovnu analitiku eksponencijalno je poraslo tokom protekle decenije (Mikalef i dr., 2019). Savremeni BA ukorenjen je u stalnom napretku sistema za podršku odlučivanju. Ovaj napredak uključuje sve snažnije mehanizme za sticanje, generisanje, asimilaciju, izbor i emitovanje znanja relevantnog za donošenje odluka. S obzirom na nasleđe podrške odlučivanju, poslovna analitika nužno učestvuje u tim mehanizmima i iskorišćava ih. Znanje koje se mora obraditi kreće se od kvalitativnog do kvantitativnog, a BA uključuje obe vrste znanja, kako je prikladno za donošenje odluke (Kohavi i dr., 2002). Razlog zašto bi neka organizacija trebala primeniti BA je u problemima koje rešava. Problemi koje BA ističe su problemi smanjenja učinkovitog upravljanja preduzećem. Shodno tome, postoji nekoliko razloga za primenu BA (Holsapple i dr., 2014):

- Ostvarivanje konkurentске prednosti;
- Podrška strateškim i taktičkim ciljevima organizacije;
- Bolji organizacioni učinak;
- Bolji ishodi odluka;
- Bolji ili informisaniji procesi odlučivanja;
- Proizvodnja znanja;
- Dobijanje vrednosti iz podataka.

Bez obzira na vrstu platforme koja se koristi u BA za rešavanje i podršku procesu donošenja odluka u svakoj kompaniji, postoje tri ključna stuba svakog BA rešenja (slika 8.1). Generalno,



BA zauzima mesto u spektru između računarske nauke/matematike/nauke o podacima (s jedne strane) i poslovanja i menadžmenta (s druge strane). Poslovna analitika zahteva i tehničko i poslovno znanje. Glavni problem u dizajniranju BA je to što granice nisu jasne (Power i dr., 2018). Shodno tome, za izvođenje BA potrebni su matematički alati za identifikaciju, izdvajanje i predstavljanje uvida na pravi način putem tabela, grafika, formula itd. Dodatno, alati za programiranje služe kao podrška ovoj vrsti aktivnosti i omogućavaju brze proračune bez grešaka, u poređenju s tradicionalnim pristupom papira i olovke. Poslednje, ali ne manje važno, potrebna je logistička ekspertiza u određenom području ili poslovnom problemu kako bi se odredili ključni uticajni faktori i povezani ekosistem.



Slika 4.48 Ključni stubovi BA u kontekstu lanca snabdevanja i logistike.

Ključni potrošač je poslovni korisnik, čiji posao, verovatno u *merchandisingu*, marketingu ili prodaji, nije direktno povezan s analitikom *per se*, ali koji obično koristi analitičke alate za poboljšanje rezultata nekog poslovnog procesa duž jedne ili više dimenzija (kao što su profit i vreme do tržišta). Poslovni korisnici ne žele imati posla s naprednim statističkim konceptima; žele jednostavne vizualizacije i rezultate relevantne za zadatak (Kohavi i dr., 2002).



8.2 Šta je R?

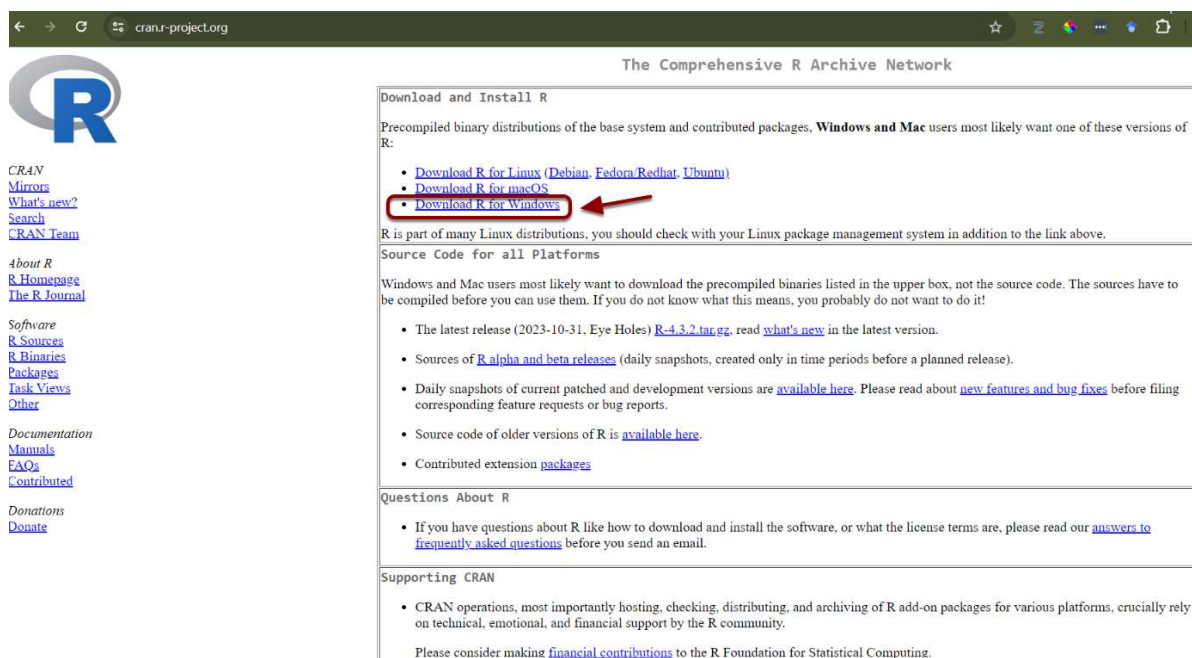
R je integrisani paket softverskih mogućnosti za manipulaciju podacima, proračun i grafički prikaz (R Core Team, 2019). Između ostalog, ima i sledeće:

- učinkovito rukovanje i skladištenje podataka,
- skup operatora za proračune na nizovima, posebno matricama,
- velika, koherentna, integrisana zbirka posrednih alata za analizu podataka,
- grafičke mogućnosti za analizu i prikaz podataka bilo direktno na računaru ili u štampanoj kopiji, te dobro razvijen, jednostavan i učinkovit programski jezik (nazvan 'S') koji uključuje uslove, petlje, korisnički definisane rekurzivne funkcije i mogućnosti unosa i izlaza (većina funkcija koje pruža sistem napisane su u S jeziku).

Glavne prednosti R-a su činjenica da je R besplatan i da postoji puno dostupne pomoći online. Prilično je sličan drugim programskim paketima kao što je MatLab (nije besplatan), ali je lakši za korišćenje od programskih jezika kao što su C++ ili Fortran (Torfs i Brauer, 2014). R je u velikoj meri sredstvo za nove metode interaktivne analize podataka. Brzo se razvijao i proširio velikom kolekcijom paketa. Međutim, većina programa napisanih u R-u u suštini su prolazni, napisani za jednu analizu podataka (R Core Team, 2019).

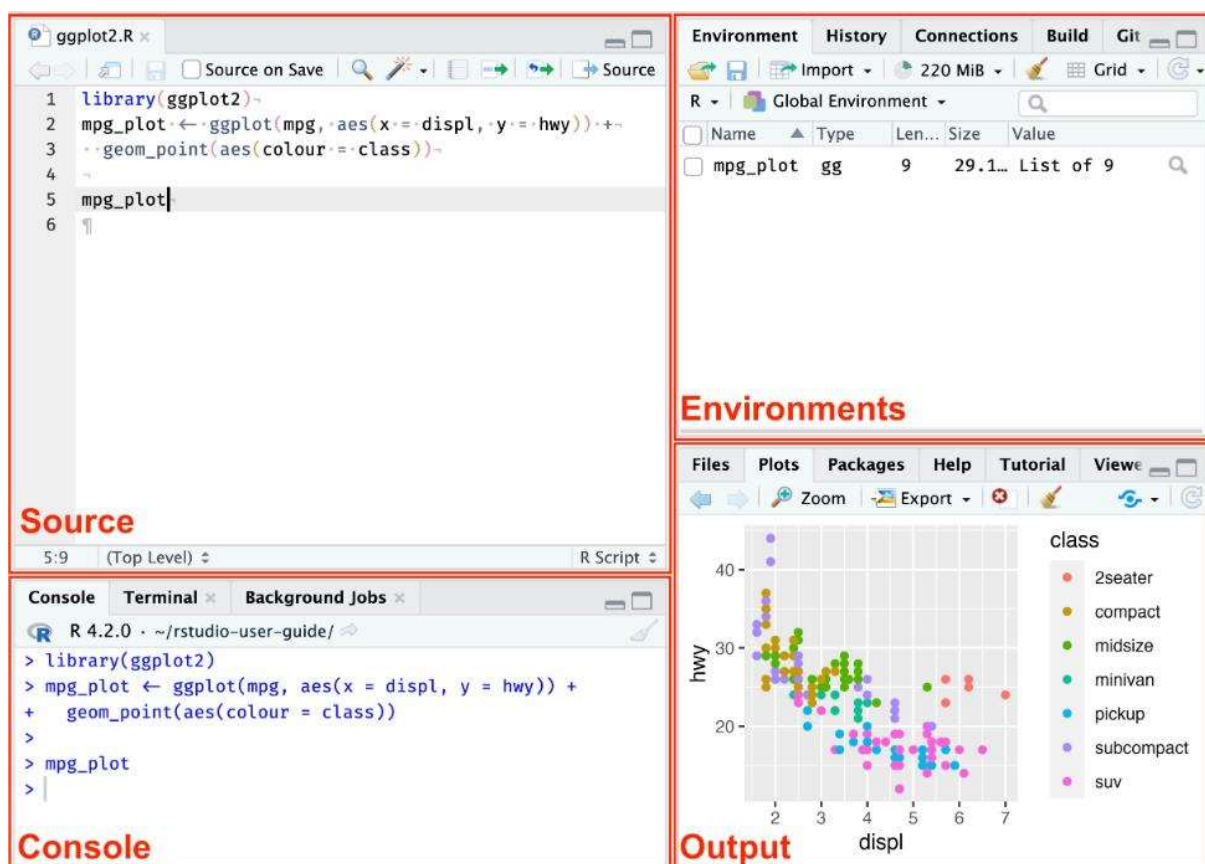
Instalacija R-a i R Studija

Da biste instalirali R, idite na cran.r-project.org i kliknite na preuzimanje R za određeni operativni sistem na vašem računaru (obično Windows) (slika 8.2).



Slika 4.49 Stranica za preuzimanje softvera R.

Ovo će preuzeti softver R, a postupak instalacije je isti kao i za bilo koji drugi softver. Kada se R softver instalira, biće bez naprednog integrisanog razvojnog okruženja (engl. *integrated development environment* - IDE) koje pomaže korisnicima u izradi različitih analiza. Iako je moguće napraviti bilo kakvu analizu samo s instaliranim R-om, poželjno ga je upariti s nekim modernim IDE-om, poput RStudio, koji je jedan od najpopularnijih IDE-a. Postupak instaliranja RStudio sličan je osnovnom R softveru. Idite na <https://posit.co/download/rstudio-desktop/>, potražite RStudio Desktop licencu otvorenog koda, preuzmite je i instalirajte. Nakon instaliranja programa R i RStudio korisnik će imati sledeći ekran korisničkog interfejsa (slika 8.3).



Slika 4.50 Korisnički interfejs i R i Rstudio (RStudio, 2024).

RStudio korisnički interfejs ima 4 primarna prozora (RStudio, 2024.):

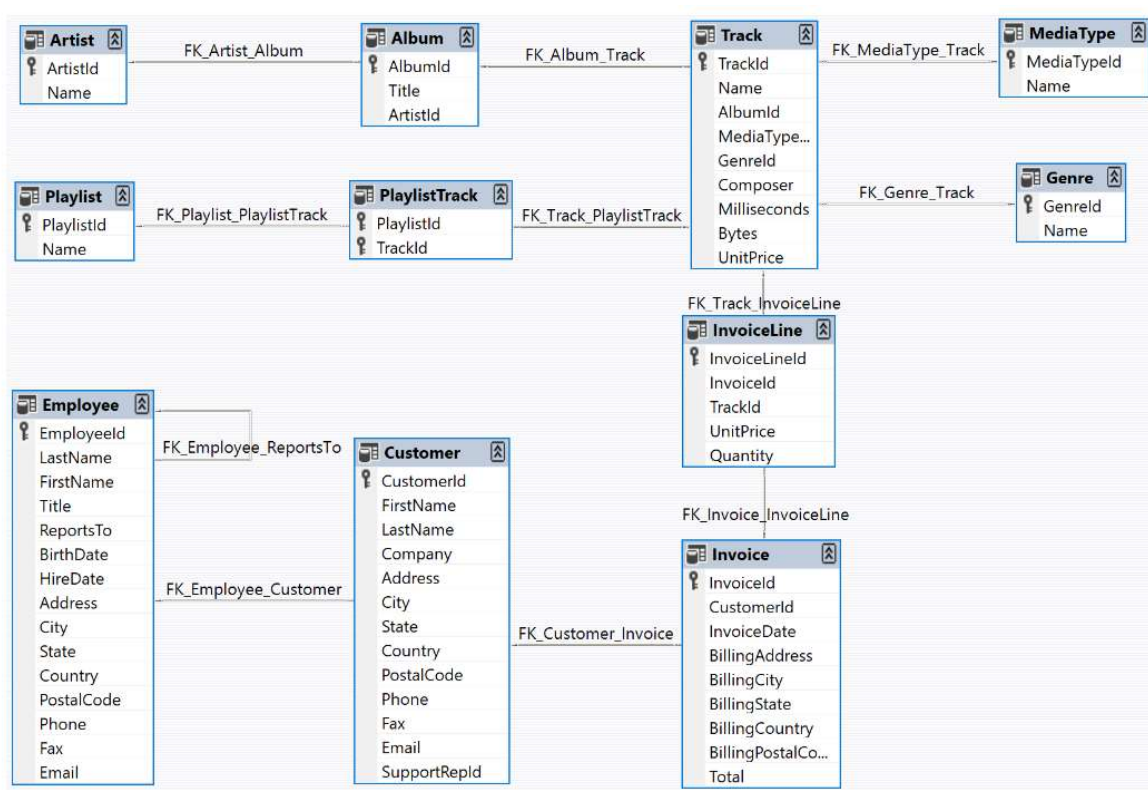
- „Source“ (srp. Izvor);
- „Console“ (srp. Konzola);
- „Environments“ (srp. Okruženje), koji sadrži kartice Okruženje, Istorija, Veze, Izrada, VCS i Vodič;
- „Output“ (srp. Rezultat), koji sadrži kartice Datoteke, Ploče, Paketi, Pomoć, Pregledač i Prezentacija.

Svaki od prozora te pododjeljak i opcije u njemu omogućavaju korisnicima izvođenje različitih operacija, kontrolu nad nekim analizama podataka ili strukturiraniji i jasniji pogled na proces analitike podataka koji je u toku.



8.3 Šta je SQL i kako je povezan s BA i R?

Chinook baza podataka je ogledna baza podataka koja se koristi za učenje i demonstraciju sistema za upravljanje bazom podataka (engl. *database management systems* - DBMS) i SQL upita. Dizajnirana je kao digitalna multimedijalna prodavnica koja koristi stvarne podatke iz iTunes knjižnice; izmišljena imena/adrese kupaca i zaposlenih; i slučajne podatke za informacije o prodaji. Baza podataka sadrži različite tabele koje predstavljaju podatke muzičke prodavnice, uključujući informacije o izvođačima, albumima, pesmama, kupcima, fakturama i slično (slika 8.4).



Slika 4.51 Model podataka baze podataka Chinook.

Slika 8.4 prikazuje model podataka koji okružuje Chinook bazu podataka s različitim tabelama podataka i njihovim ključevima (jedinstveni identifikatori) i zajedničkim tabelama poput tabele PlaylistTrack. Tabele prenose različite informacije o određenoj digitalnoj prodavnici (tabela 8.1).



Tabela 4.3 Informacije sadržane u svakoj od tabela baze podataka Chinook.

Naziv tabele	Opis
Umetnik	Sadrži informacije o muzičkim umetnicima.
Album	Sadrži informacije o muzičkim albumima, od kojih je svaki povezan s izvođačem.
Zapis	Sadrži informacije o pojedinačnim muzičkim zapisima, uključujući reference na albume, vrste medija i žanrove.
Žanr	Sadrži informacije o muzičkim žanrovima.
Vrsta medija	Sadrži informacije o različitim vrstama medija (npr. audio, video).
Kupac	Sadrži podatke o kupcima, uključujući podatke za kontakt i podatke o predstavniku podrške.
Zaposleni	Sadrži informacije o zaposlenima, uključujući njihove uloge, veze s izveštajima i podatke za kontakt.
Fakture	Sadrži informacije o fakturama, uključujući pojedinosti o kupcima, podatke o naplati i ukupne iznose.
Linija fakture	Sadrži detaljne informacije o svakoj stavci na fakturi, uključujući reference na tragove i količine.
Popis pesama	Sadrži informacije o popisima za reprodukciju.
Zapis popisa pesama	Povezuje pesme s popisima za reprodukciju, pokazujući koje su pesme uključene u koje popise za reprodukciju.

8.4 Kako su povezani poslovna analitika, SQL i R?

Veza između BA, SQL i R je prirodna jer bi svi poslovni podaci trebali da se skladište u SQL bazama podataka. Ovo je još uvek idealistički cilj budući da još uvek postoji loše upravljanje podacima u nekim delovima malih i srednjih preduzeća koja još uvek ne razumeju u potpunosti snagu podataka. U velikim kompanijama to je davno prepoznato i podaci su pravilno strukturirani u bazama podataka (SQL ili nekoj drugoj, ali najčešće u SQL-u). S druge strane, analizu podataka moguće je izvršiti u SQL-u, ali je u tu svrhu bolje koristiti statistički orijentisani softver, gde je u fokusu R, kao jedna od najpopularnijih statističkih platformi za analizu podataka.

Shodno tome, SQL i R mogu se smatrati savršenim alatom za saradnju kada je problem u pitanju iz područja BA. Postoji nekoliko glavnih razloga, a jedan od njih je taj što se BA podaci svakodnevno menjaju i ažuriraju u skladu s realnim tržištem i aktivnostima kompanije: prodaja, zaposleni, prihodi itd. SQL baze podataka savršene su za beleženje tih promena i ažuriranje postojećih podataka, dok su R skripte vrlo dobre u automatizaciji zadataka kao i u dizajniranju



novih paketa za analizu datih podataka. Razlog za to je što je R više građen oko koncepta analize podataka, nego na opštem programiranju kao što je Python, na primer.

Upit SQL bazi podataka s R

R i SQL baze podataka imaju prirodnu vezu budući da je R uglavnom izgrađen za statističku analizu podataka, a većina transakcijskih podataka nalazi se u bazama podataka. "Način" na koji R radi za upravljanje manipulacijom podataka iz SQL baza podataka je preko DBI i RSQLite programskih paketa. DBI paket pruža standardizovani interfejs za interakciju s različitim DBMS-ovima, dopuštajući korisnicima povezivanje, postavljanje upita i dosledno upravljanje transakcijama u različitim bazama podataka. Paket RSQLite, koji se pridržava DBI interfejsa, posebno olakšava interakciju s bazama podataka SQLite, omogućavajući korisnicima izvršavanje SQL upita, preuzimanje podataka i izvođenje operacija baze podataka direktno iz R-a. Zajedno, ovi paketi pojednostavljuju proces rada s bazama podataka u R-u, nudeći kohezivan i učinkovit tok rada.

Kako bismo učinkovito demonstrirali izvođenje SQL operacije iz R-a i generisanje željenih uvida iz podataka u vezi s postojećim problemom, dali smo nekoliko isečaka koda na slikama 8.5 i 8.6.



```
---
title: "BUSINESS ANALYTICS FOUNDATIONS INCLUDING THE R AND SQL"
format: html
editor: visual
---

# R & SQL

## Loading libraries

```{r setup, warning=FALSE, message=FALSE}
library(DBI)
library(RSQLite)
```

## Connect to the Chinook SQLite database

```{r}
con <- dbConnect(RSQLite::SQLite(), dbname = "Chinook_Sqlite.sqlite")
```

## List all tables in the database

```{r}
tables <- dbListTables(con)
print(tables)
```
```

Slika 4.52 Isečak koda za uspostavljanje veze između SQL-a i R-a i istraživanje tabela podataka sadržanih u SQL-u.

Prvi korak u postavljanju upita SQL bazi podataka putem R-a je uspostavljanje veze (slika 8.5). Slika prikazuje korišćenje DBI i RSQLite paketa koji omogućavaju uspostavljanje veze putem `dbConnect()` funkcije. Rezultat veze i tabele podataka koji se otkrivaju putem gore navedene veze zatim se izvoze putem funkcija `dbListTables()` koje ispisuju popis svih podataka pronađenih putem veze: Album, Izvođač, Kupac, Zaposleni, Žanr, Fature, Linija fakture, Vrsta medija, Popis pesama, Zapis popisa pesama, Zapis.

Nakon što se veza uspostavi, postoji niz mogućih analiza koje se mogu sprovesti, zavisno od poslovnog cilja i budućoj upotrebi datih rezultata. Ovde ćemo, zbog ograničenja prostora, pokazati samo delić moguće analize podataka, s malim isečkom koda i skupom pravila koda potrebnih za izdvajanje informacija iz SQL-a. Kod vrši upite bazi podataka putem R-a i prikazuje najprodavanije albume, njihove autore i prodani broj (slika 8.6). Tabela 8.2 predstavlja rezultate upita podataka putem isečka koda na slici 8.6.



```
29 ## Choose a Album table from the database
30 ```{r}
31 query_album <- "SELECT * FROM Album LIMIT 10"
32 data_album <- dbGetQuery(con, query_album)
33 print(data_album)
34 ```
35
36 ## Query to get album details along with artist names
37 ```{r}
38 query_album_artist <- "
39 SELECT Album.AlbumId, Album.Title AS AlbumTitle, Artist.Name AS ArtistName
40 FROM Album
41 JOIN Artist ON Album.ArtistId = Artist.ArtistId
42 LIMIT 10"
43
44 data_album_artist <- dbGetQuery(con, query_album_artist)
45 print(data_album_artist)
46 ```
47
48 ## Query to get the top-selling albums along with artist names
49 ```{r}
50 query_top_selling_albums <- "
51 SELECT
52     Album.Title AS AlbumTitle,
53     Artist.Name AS ArtistName,
54     SUM(InvoiceLine.Quantity) AS TotalQuantitySold
55 FROM
56     InvoiceLine
57 JOIN
58     Track ON InvoiceLine.TrackId = Track.TrackId
59 JOIN
60     Album ON Track.AlbumId = Album.AlbumId
61 JOIN
62     Artist ON Album.ArtistId = Artist.ArtistId
63 GROUP BY
64     Album.AlbumId, Album.Title, Artist.Name
65 ORDER BY
66     TotalQuantitySold DESC
67 LIMIT 10"
68
69 # Execute the query
70 top_selling_albums <- dbGetQuery(con, query_top_selling_albums)
71 knitr::kable(top_selling_albums)
72 ```
```

Slika 4.53Isečak koda za postavljanje upita SQL-u putem R-a i određivanje 10 najprodavanijih albuma.

Tabela 4.410 najprodavanijih albuma u digitalnoj prodavnici Chinook.

| Naslov albuma | Ime umetnika | Prodana količina |
|-------------------|------------------------------|------------------|
| Minha Historia | Chico Buarque | 27 |
| Greatest Hits | Lenny Kravitz | 26 |
| Unplugged | Eric Clapton | 25 |
| Acústico | Titãs | 22 |
| Greatest Kiss | Kiss | 20 |
| Prenda Minha | Caetano Veloso | 19 |
| Chronicle, Vol. 2 | Creedence Clearwater Revival | 19 |



| Naslov albuma | Ime umetnika | Prodata količina |
|--|------------------------------|------------------|
| My Generation - The Very Best Of The Who | The Who | 19 |
| International Superhits | Green Day | 18 |
| Chronicle, Vol. 1 | Creedence Clearwater Revival | 18 |

LITERATURA 8. POGLAVLJA

- Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, 64, 130-141.
- Kohavi, R., Rothleder, N. J., & Simoudis, E. (2002). Emerging trends in business analytics. *Communications of the ACM*, 45(8), 45-48.
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data and business analytics: A research agenda for realizing business value. *Information & Management*. <https://doi.org/10.1016/j.im.2019.103237>
- Power, D. J., Heavin, C., McDermott, J., & Daly, M. (2018). Defining business analytics: An empirical approach. *Journal of Decision Systems*, 27(1), 40–53. <https://doi.org/10.1080/2573234X.2018.1507605>
- R Core Team. (2019). An introduction to R: Notes on R, a programming environment for data analysis and graphics. The R Foundation.
- RStudio. (2024). RStudio IDE cheatsheet: UI panes. Posit. <https://docs.posit.co/ide/user/ide/guide/ui/ui-panes.html>
- Torfs, P., & Brauer, C. (2014). A (very) short introduction to R. Hydrology and Quantitative Water Management Group, Wageningen University, The Netherlands, 1-12.