



8. Temelji poslovne analitike uključujući R i SQL

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

Na ova i slična pitanja pokušat ćemo dati odgovore u sljedećem poglavlju.

8.1 Što je poslovna analitika?

BA predstavlja holistički pristup analizi podataka i poslovnom odlučivanju. To je okruženje vođeno podacima s ciljem poboljšanja poslovnih performansi tvrtke pružanjem temelja za informiranije donošenje odluka. To je sustavni proces razmišljanja koji primjenjuje kvalitativne, kvantitativne i statističke računalne alate i metode za analizu podataka, stjecanje uvida, informiranje i podršku donošenju odluka. Svaka određena analiza može koristiti različite tehnike uključujući dijagnostičke, prediktivne, preskriptivne i optimizacijske modele (Power i dr., 2018). Mikalef i dr., (2019) daju plan za akademsko istraživanje i praktičnu primjenu, ističući transformativni potencijal analitike kada se pravilno integrira u organizacijske procese. U skladu s tim, autori navode da BA zahtijeva od organizacija da radikalno redizajniraju način na koji se takvim inicijativama pristupa, kako se dizajniraju i usavršavaju, kako se planiranje resursa i orkestracija izvršava i strateški usklađuje, kao i da ponovno vrednuju svoje očekivane rezultate izvedbe, 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. Opći cilj je poslovna učinkovitost kroz 'vertikalizaciju', upotrebljivost i integraciju s operativnim sustavima (Kohavi i dr., 2002). BA ima mnogo područja primjene i povezanih izvedenica: financijska analitika, analitika opskrbnog lanca, 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** – gledaju postojeće podatke i daju sažetak statistike i osnovnu vizualizaciju.
- **Prediktivne** – koriste postojeće podatke za procjenu najvjerojatnijih budućih scenarija.
- **Preskriptivne** – automatski obrađuju veliku količinu podataka (engl. *big data*), poslovna pravila, tržišne uvjete itd. Ove platforme koriste metode strojnog učenja i umjetne inteligencije. Cilj je potpuno automatizirano donošenje odluka o tome koje akcije tvrtka treba poduzeti s obzirom na trenutnu situaciju kako bi postigla željene poslovne ciljeve.

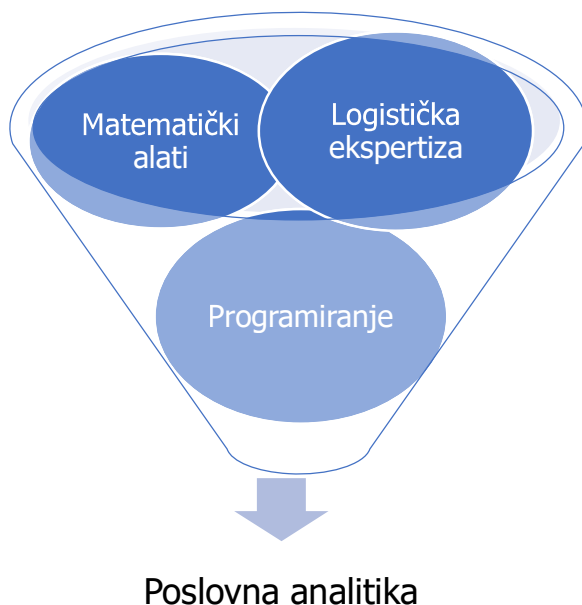
Zanimanje za *big data* i poslovnu analitiku eksponencijalno je poraslo tijekom proteklog desetljeća (Mikalef i dr., 2019). Suvremeni BA ukorijenjen je u stalnom napretku sustava za podršku odlučivanju. Ovaj napredak uključuje sve snažnije mehanizme za stjecanje, generiranje, asimilaciju, odabir i emitiranje znanja relevantnog za donošenje odluka. S obzirom na nasljeđe podrške odlučivanju, poslovna analitika nužno sudjeluje u tim mehanizmima i iskorištava ih. Znanje koje se mora obraditi kreće se od kvalitativnog do kvantitativnog, a BA se bavi radom na obje vrste znanja, kako je prikladno za donošenje odluke (Kohavi i dr., 2002). Razlog zašto bi neka organizacija trebala primijeniti BA je u problemima koje rješava. Problemi koje BA ističe su problemi smanjenja učinkovitog upravljanja poduzećem. Sukladno tome, postoji nekoliko razloga za primjenu BA (Holsapple i dr., 2014):

- Ostvarivanje konkurentske prednosti
- Podrška strateškim i taktičkim ciljevima organizacije
- Bolji organizacijski učinak
- Bolji ishodi odluka
- Bolji ili informiraniji procesi odlučivanja
- Proizvodnja znanja
- Dobivanje vrijednosti iz podataka

Bez obzira na vrstu platforme koja se koristi u BA za rješavanje i podršku procesu donošenja odluka u svakoj tvrtki, postoje tri ključna stupa svakog BA rješenja (slika 8.1). Općenito, BA zauzima mjesto u spektru između računalnih znanosti/matematike/podatkovnih znanosti (s



jedne strane) i poslovanja i menadžmenta (s druge strane). Poslovna analitika zahtijeva i tehničko i poslovno znanje. Glavni problem u dizajniranju BA je to što granice nisu jasne (Power i dr., 2018). Sukladno tome, za izvođenje BA potrebni su matematički alati za identifikaciju, izdvajanje i predstavljanje uvida na pravi način putem tablica, grafika, formula itd. Dodatno, alati za programiranje služe kao potpora ovoj vrsti aktivnosti i omogućuju brze izračuna bez grešaka, u usporedbi s tradicionalnim pristupom papira i olovke. Posljednje, ali ne manje važno, potrebna je logistička ekspertiza u određenom području ili poslovnom problemu kako bi se odredili ključni utjecajni čimbenici i povezani ekosustav.



Slika 4.48 Ključni stupovi BA u kontekstu opskrbnog lanca i logistike.

Ključni potrošač je poslovni korisnik, čiji posao, vjerojatno u *merchandisingu*, marketingu ili prodaji, nije izravno 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 vrijeme 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 Što je R?

R je integrirani paket softverskih mogućnosti za manipulaciju podacima, izračun i grafički prikaz (R Core Team, 2019). Između ostalog, ima i sljedeće:

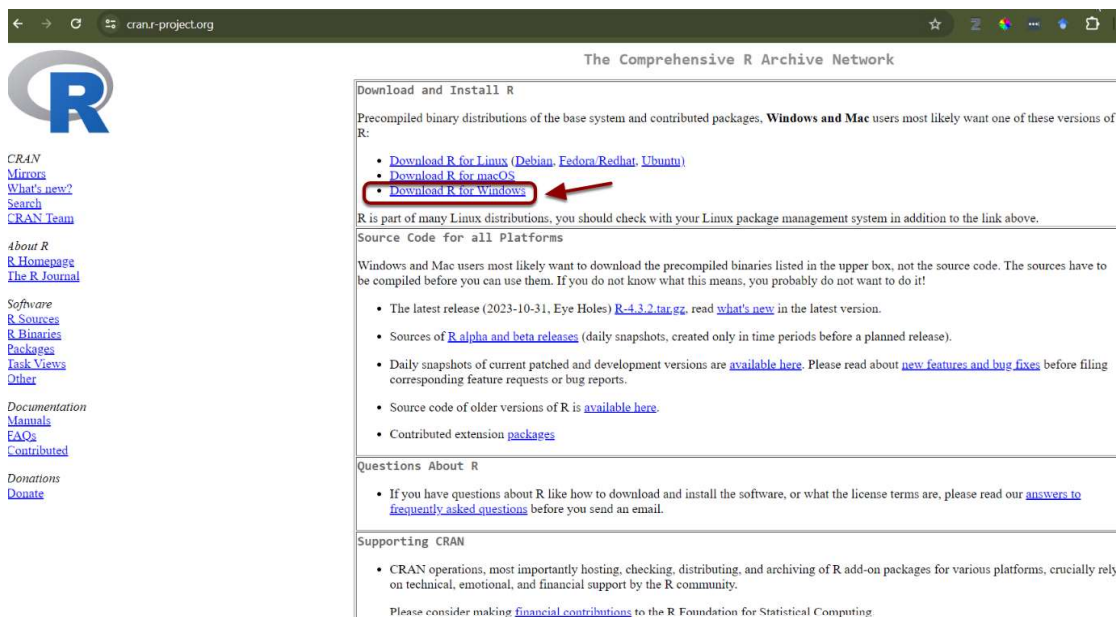


- učinkovito rukovanje i skladištenje podataka,
- skup operatora za izračune na nizovima, posebno matricama,
- velika, koherentna, integrirana zbirka posrednih alata za analizu podataka,
- grafičke mogućnosti za analizu i prikaz podataka bilo izravno na računalu ili u tiskanoj kopiji, te dobro razvijen, jednostavan i učinkovit programski jezik (nazvan 'S') koji uključuje uvjete, petlje, korisnički definirane rekurzivne funkcije i mogućnosti unosa i izlaza (većina funkcija koje pruža sustav same su napisane 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 korištenje od programskih jezika kao što su C++ ili Fortran (Torfs i Brauer, 2014). R je u velikoj mjeri sredstvo za nove metode interaktivne analize podataka. Brzo se razvijao i proširio velikom kolekcijom paketa. Međutim, većina programa napisanih u R-u u biti su prolazni, napisani za jednu analizu podataka (R Core Team, 2019).

Instalacija R-a i R Studija

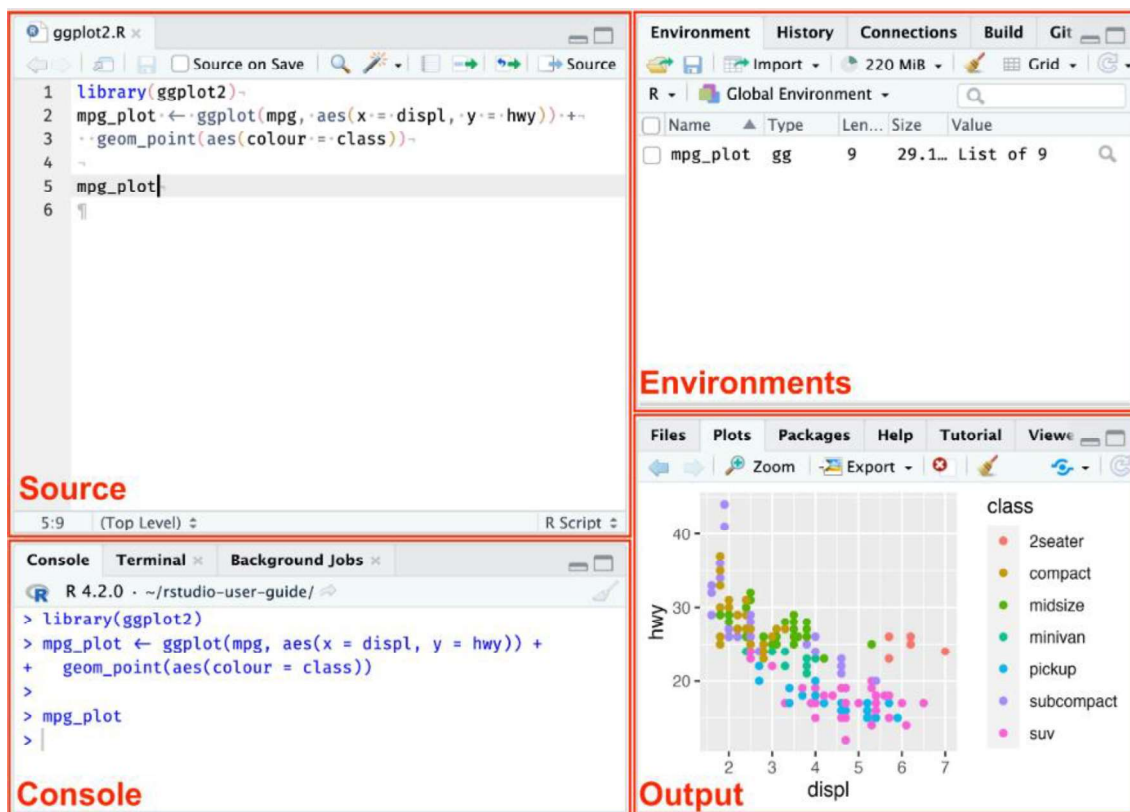
Da biste instalirali R, idite na cran.r-project.org i kliknite na preuzimanje R za određeni operativni sustav na vašem računalu (obično Windows) (slika 8.2).



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, bit će bez naprednog integriranog 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 RStudija, koji je jedan od najpopularnijih IDE-a. Postupak instaliranja RStudija 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 sljedeći zaslon korisničkog sučelja (slika 8.3).



Slika 4.50Korisničko sučelje i R i Rstudio (RStudio, 2024).

RStudio korisničko sučelje ima 4 primarna prozora (RStudio, 2024.):

- „Source“ (hrv. Izvor);
- „Console“ (hrv. Konzola);
- „Environments“ (hrv. Okruženje), koji sadrži kartice Okruženje, Povijest, Veze, Izrada, VCS i Vodič;

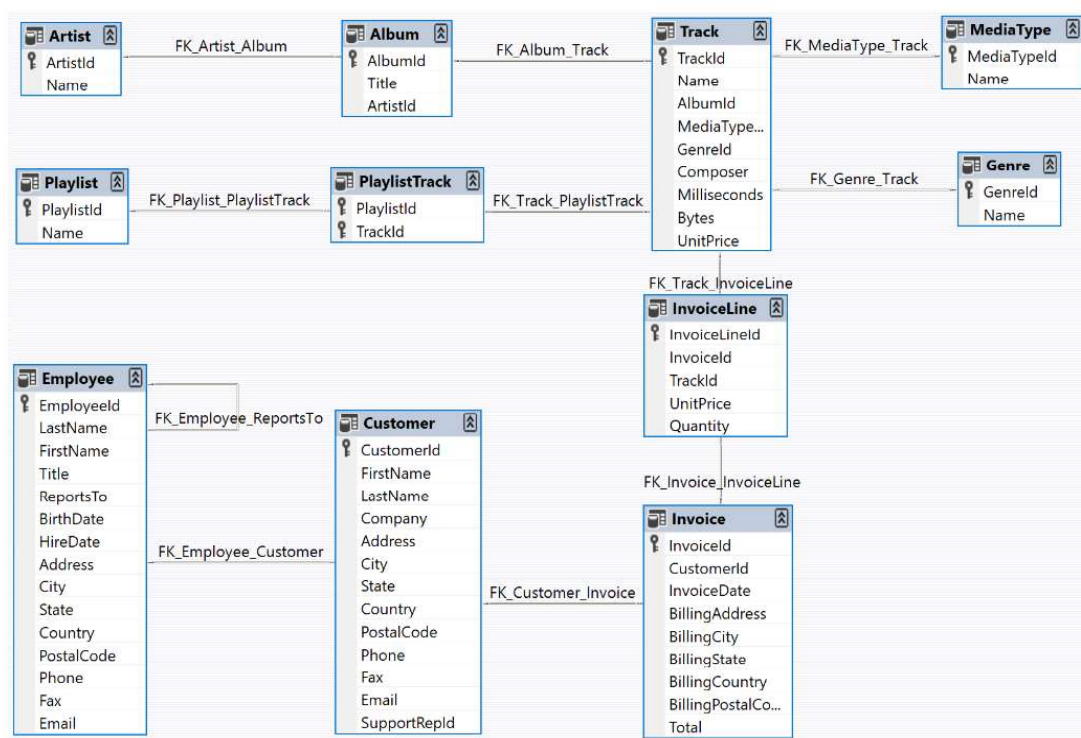


- „Ouput“ (hrv. Rezultat), koji sadrži kartice Datoteke, Ploče, Paketi, Pomoć, Preglednik i Prezentacija.

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

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

Chinook baza podataka je ogledna baza podataka koja se koristi za učenje i demonstraciju sustava za upravljanje bazom podataka (engl. *database management systems* - DBMS) i SQL upita. Dizajniran je kao digitalna multimedijaska prodavaonica koja koristi stvarne podatke iz iTunes knjižnice; izmišljena imena/adrese kupaca i zaposlenika; i slučajni podaci za informacije o prodaji. Baza podataka sadrži različite tablice koje predstavljaju podatke glazbene prodavaonice, uključujući informacije o izvođačima, albumima, pjesmama, kupcima, fakturama i više (slika 8.4).



Slika 4.51 Model podataka baze podataka Chinook.



Slika 8.4 prikazuje model podataka koji okružuje Chinook bazu podataka s različitim podatkovnim tablicama i njihovim ključevima (jedinствени identifikatori) i zajedničkim tablicama poput tablice PlaylistTrack. Tablice prenose različite informacije o određenoj digitalnoj prodavaonici (tablica 8.1).

Tablica 4.3 Informacije sadržane u svakoj od tablica baze podataka Chinook.

Naziv tablice	Opis
Umjetnik	Sadrži informacije o glazbenim umjetnicima.
Album	Sadrži informacije o glazbenim albumima, od kojih je svaki povezan s izvođačem.
Zapis	Sadrži informacije o pojedinačnim glazbenim zapisima, uključujući reference na albume, vrste medija i žanrove.
Žanr	Sadrži informacije o glazbenim 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.
Zaposlenik	Sadrži informacije o zaposlenicima, uključujući njihove uloge, odnose s izvješćima 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 pjesama	Sadrži informacije o popisima za reprodukciju.
Zapis popisa pjesama	Povezuje pjesme s popisima za reprodukciju, pokazujući koje su pjesme uključene u koje popise za reprodukciju.

8.4 Kako su poslovna analitika, SQL i R povezani?

Veza između BA, SQL i R je prirodna jer bi svi poslovni podaci trebali biti pohranjeni u SQL bazama podataka. Ovo je još uvijek idealistički cilj budući da još uvijek postoji loše upravljanje podacima u nekim dijelovima malih i srednjih poduzeća koja još uvijek ne razumiju u potpunosti snagu podataka. U velikim tvrtkama to je davno prepoznato i podaci su pravilno strukturirani u bazama podataka (SQL ili neki drugi, ali najčešće u SQL-u). S druge strane, analizu podataka moguće je izvršiti u SQL-u, ali je u tu svrhu bolje koristiti statistički orijentirani softver, gdje je u fokusu R, kao jedna od najpopularnijih statističkih platformi za analizu podataka.



Sukladno tome, SQL i R mogu se smatrati savršenim alatom za suradnju kada je problem u pitanju iz područja BA. Postoji nekoliko glavnih razloga, a jedan od njih je taj što se BA podaci svakodnevno mijenjaju i ažuriraju u skladu s realnim tržištem i aktivnostima tvrtke: prodaja, zaposlenici, prihodi itd. SQL baze podataka savršene su za bilježenje tih promjena i ažuriranje postojećih podataka, dok su R skripte vrlo dobre u automatizaciji zadataka kao i u dizajniranju novih paketa za analizu danih podataka. Razlog za to je što je R više izgrađen oko koncepta analize podataka, nego na općem programiranju kao što je Python, na primjer.

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 standardizirano sučelje za interakciju s različitim DBMS-ovima, dopuštajući korisnicima povezivanje, postavljanje upita i dosljedno upravljanje transakcijama u različitim bazama podataka. Paket RSQLite, koji se pridržava DBI sučelja, posebno olakšava interakciju s bazama podataka SQLite, omogućujući korisnicima izvršavanje SQL upita, dohvaćanje podataka i izvođenje operacija baze podataka izravno iz R-a. Zajedno, ovi paketi pojednostavljuju proces rada s bazama podataka u R-u, nudeći kohezivan i učinkovit tijek rada.

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



```
---
title: "BUSINES ANALYTICS FOUNDATINS 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 Isječak koda za uspostavljanje veze između SQL-a i R-a i istraživanje podatkovnih tablica sadržanih u SQL-u.

Prvi korak u postavljanju upita SQL bazi podataka putem R-a je uspostavljanje veze (slika 8.5). Slika prikazuje korištenje DBI i RSQLite paketa koji omogućuju uspostavljanje veze putem dbConect() funkcije. Rezultat veze i podatkovne tablice koje se otkrivaju putem gore navedene veze zatim se izvoze putem funkcija dbListTables() koje ispisuju popis svih podataka pronađenih putem veze: Album, Izvođač, Kupac, Zaposlenik, Žanr, Fature, Linija fature, Vrsta medija, Popis pjesama, Zapis popisa pjesama, Zapis.

Nakon što se veza uspostavi, postoji niz mogućih analiza koje se mogu provesti, ovisno o poslovnom cilju i budućoj upotrebi danih rezultata. Ovdje ćemo, zbog ograničenja prostora, pokazati samo djelić moguće analize podataka, s malim isječkom 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). Tablica 8.2 predstavlja rezultate upita podataka putem isječka 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 4.53Isječak koda za postavljanje upita SQL-u putem R-a i određivanje 10 najprodavanijih albuma.

Tablica 4.410 najprodavanijih albuma u digitalnoj prodavaonici Chinook.

| Naslov albuma | Ime umjetnika | 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 umjetnika | Prodana 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.