



8. Osnove poslovne analitike, vključno z R in SQL

Kaj je poslovna analitika (BA)? Katere probleme rešuje in katera orodja uporablja? Kaj sta R in SQL? Kako je BA povezana z R in SQL? Ali obstajajo primeri dobrih praks, kjer se ti programi uporabljajo za reševanje poslovnih problemov v logistiki?

Na ta in podobna vprašanja bomo poskušali odgovoriti v naslednjem poglavju.

8.1 Kaj je poslovna analitika?

BA predstavlja celovit pristop k analizi podatkov in poslovnemu odločanju. Gre za okolje, ki temelji na podatkih, njegov cilj pa je izboljšati poslovno uspešnost podjetja z zagotavljanjem podlage za bolj premišljeno odločanje. Gre za sistematičen proces razmišljanja, ki uporablja kvalitativna, kvantitativna in statistična računalniška orodja in metode za analizo podatkov, pridobivanje vpogleda, informiranje in podporo odločanju. Vsaka posamezna analiza lahko uporablja različne tehnike, vključno z diagnostičnimi, napovednimi, preskriptivnimi in optimizacijskimi modeli (Power et al., 2018). Tako po Mikalef et al. (2019) BA zagotavlja načrt tako za akademsko raziskovanje kot za praktično izvajanje, pri čemer poudarja transformativni potencial analitike, če je ustrezno vključena v organizacijske procese. V skladu s tem avtorji navajajo, da BA od organizacij zahteva, da korenito preoblikujejo način pristopa, načrtovanja in izpopolnjevanja takšnih pobud, način načrtovanja in upravljanja virov ter strateško uskladitev, pa tudi ponovno ovrednotijo pričakovane rezultate delovanja, njihovo povezanost s strateškimi cilji in posledično razvijejo ustrezne KPI (Mikalef et al., 2019).

Glavna naloga BA je zagotoviti prenos znanja, da se zagotovi skladna povezava med neobdelanimi podatki in poslovnimi odločitvami. Splošni cilj je poslovna učinkovitost z "vertikalizacijo", uporabnostjo in integracijo z operativnimi sistemi (Kohavi, Rothleder in Simoudis, 2002). BA ima številna področja uporabe in povezane izpeljanke: Finančna analitika, analitika oskrbovalne verige, analitika kriz, analitika znanja, analitika trženja, analitika strank, analitika storitev, analitika človeških virov, analitika talentov, analitika procesov, analitika tveganja (Holsapple, Lee-Post in Pakath, 2014).

Obstajajo tri vrste platform za poslovno analitiko:



- **Opisni** - pregledajo obstoječe podatke ter zagotovijo zbirne statistične podatke in osnovno vizualizacijo.
- **Prediktivni** - na podlagi obstoječih podatkov ocenjujejo najverjetnejše scenarije za prihodnost.
- **Preskriptivni** - samodejno obdelujejo velike količine podatkov, poslovna pravila, tržne razmere itd. Te platforme uporabljajo metode strojnega učenja in umetne inteligence. Cilj je popolnoma avtomatizirano odločanje o tem, katere ukrepe naj podjetje sprejme glede na trenutne razmere, da bo doseglo želene poslovne cilje.

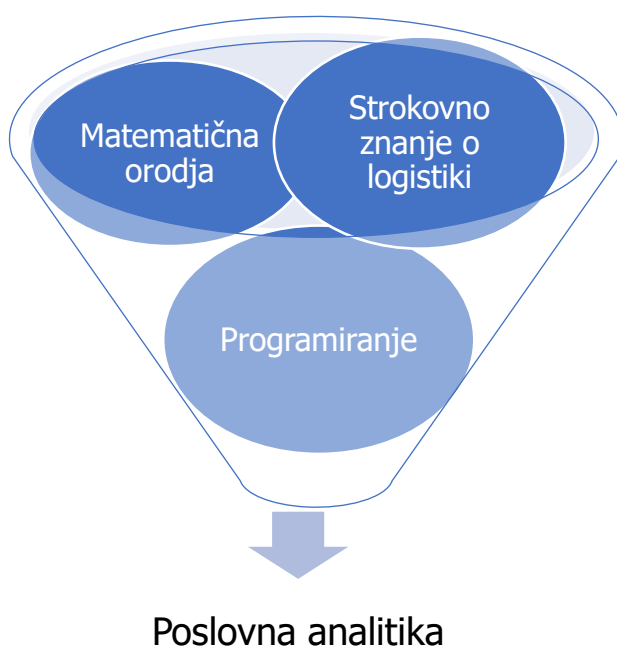
Zanimanje za velike podatke in poslovno analitiko se je v zadnjem desetletju eksponentno povečalo (Mikalef et al., 2019). Sodobni BA korenini v nenehnem napredku sistemov za podporo odločanju. Ti napredki vključujejo vse zmogljivejše mehanizme za pridobivanje, ustvarjanje, asimilacijo, izbiro in oddajanje znanja, pomembnega za sprejemanje odločitev. Glede na dediščino podpore odločanju je poslovna analitika nujno del teh mehanizmov in jih izkorišča. Znanje, ki ga je treba obdelati, sega od kvalitativnega do kvantitativnega, BA pa se ukvarja z delovanjem na obeh vrstah znanja, kot je primerno za zadevno odločitev (Kohavi, Rothleder in Simoudis, 2002). Razlog, zakaj naj bi nekatere organizacije uporabljale BA, je v problemih, ki jih rešuje. Težave, ki jih poudarja BA, so omejene težave za učinkovito upravljanje podjetja. V skladu s tem obstaja več razlogov za uporabo BA (Holsapple, Lee-Post in Pakath, 2014):

- Doseganje konkurenčne prednosti
- Podpiranje strateških in taktičnih ciljev organizacije
- Boljša organizacijska uspešnost
- Boljši rezultati odločitev
- Boljši ali bolj informirani procesi odločanja.
- Pridobivanje znanja
- Pridobivanje vrednosti iz podatkov

Ne glede na vrsto platforme, ki se uporablja v BA za reševanje in podporo procesa odločanja v posameznem podjetju, obstajajo trije ključni stebri vsake rešitve BA (slika 8.1). Na splošno BA zavzema mesto v spektru med računalništvom/matematiko/podatkovno znanostjo (na eni strani) ter poslovanjem in upravljanjem (na drugi strani). Poslovna analitika zahteva tako tehnično kot poslovno znanje. Glavna težava pri oblikovanju BA je, da meje niso jasno razmejene (Power et al.,



2018). V skladu s tem so za izvedbo BA potrebna matematična orodja, ki omogočajo prepoznavanje, pridobivanje in predstavitev vpogledov na ustrezen način prek tabel, grafikonov, formul itd. Poleg tega programska orodja služijo kot podpora tovrstnim dejavnostim ter omogočajo hitre in brezhibne izračune v primerjavi s tradicionalnim pristopom s papirjem in pisalom. Ne nazadnje je potrebno tudi strokovno znanje s področja logistike na določenem področju ali pri določenem poslovnem problemu, da se natančno določijo ključni vplivni dejavniki in z njimi povezan ekosistem.



Slika 8.1 Ključni stebri BA v okviru oskrbovalne verige in logistike.

Ključni uporabnik je poslovni uporabnik, čigar delo, po možnosti na področju prodaje, trženja ali prodaje, ni neposredno povezano z analitiko kot tako, vendar običajno uporablja analitična orodja za izboljšanje rezultatov nekega poslovnega procesa po eni ali več dimenzijah (kot sta dobiček in čas za vstop na trg). Poslovni uporabniki se ne želijo ukvarjati z naprednimi statističnimi koncepti; želijo preproste vizualizacije in rezultate, ki so pomembni za nalogo (Kohavi, Rothleder in Simoudis, 2002).

8.2 Kaj je R?

R je integriran paket programske opreme za manipulacijo s podatki, izračunavanje in grafično prikaz (R Core Team, 2019). Med drugim ima:

- učinkovito orodje za obdelavo in shranjevanje podatkov,

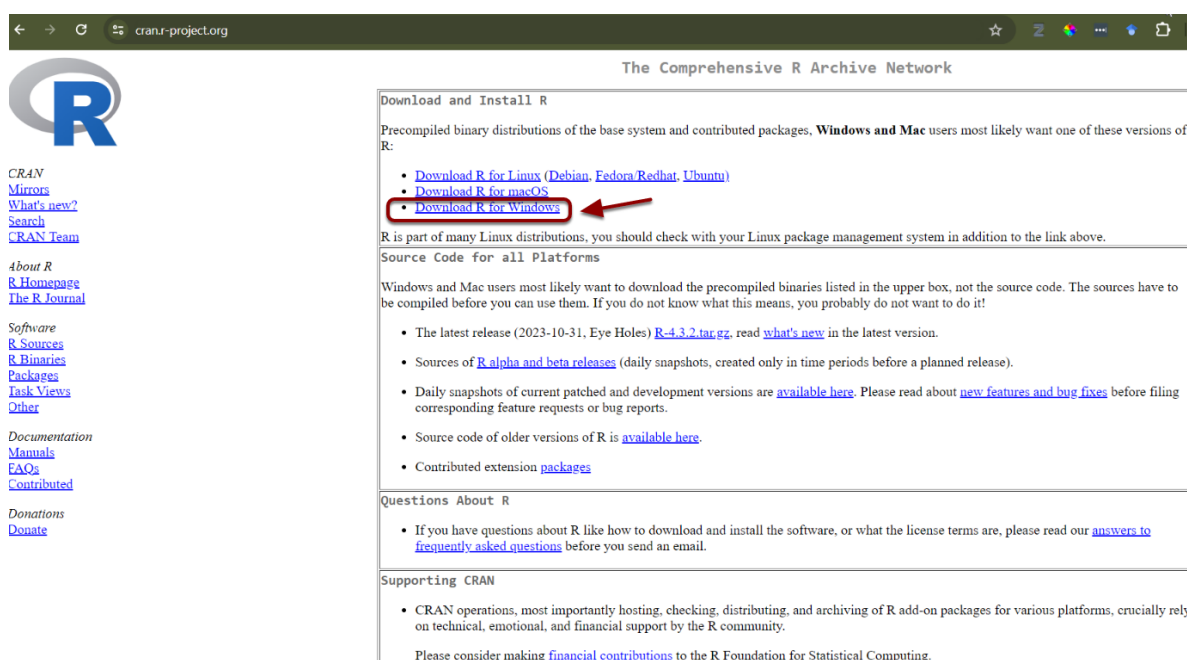


- nabor operatorjev za izračune na poljih, zlasti matrikah,
- velika, skladna in integrirana zbirka vmesnih orodij za analizo podatkov,
- Grafične možnosti za analizo in prikaz podatkov neposredno na računalniku ali na papirju ter dobro razvit, preprost in učinkovit programski jezik (imenovan "S"), ki vključuje pogoje, zanke, rekurzivne funkcije, ki jih določa uporabnik, ter vhodne in izhodne možnosti. (Večina funkcij, ki jih ponuja sistem, je dejansko napisana v jeziku S.)

Glavni prednosti programa R sta, da je brezplačen in da je na voljo veliko pomoči na spletu. Je precej podoben drugim programskim paketom, kot je MatLab (ni brezplačen), vendar je uporabniku prijaznejši od programskih jezikov, kot sta C++ ali Fortran (Torfs in Brauer, 2014). R je v veliki meri sredstvo za nov razvoj metod interaktivne analize podatkov. Hitro se je razvijal in bil razširjen z veliko zbirko paketov. Vendar je večina programov, napisanih v R, v bistvu efemernih, napisanih za eno samo analizo podatkov (R Core Team, 2019).

Namestitev programov R in R Studio

Če želite namestiti R, pojdite na stran [r-project.org](https://cran.r-project.org) in kliknite spustni gumb R za določen operacijski sistem v računalniku (običajno Windows) (slika 8.2).

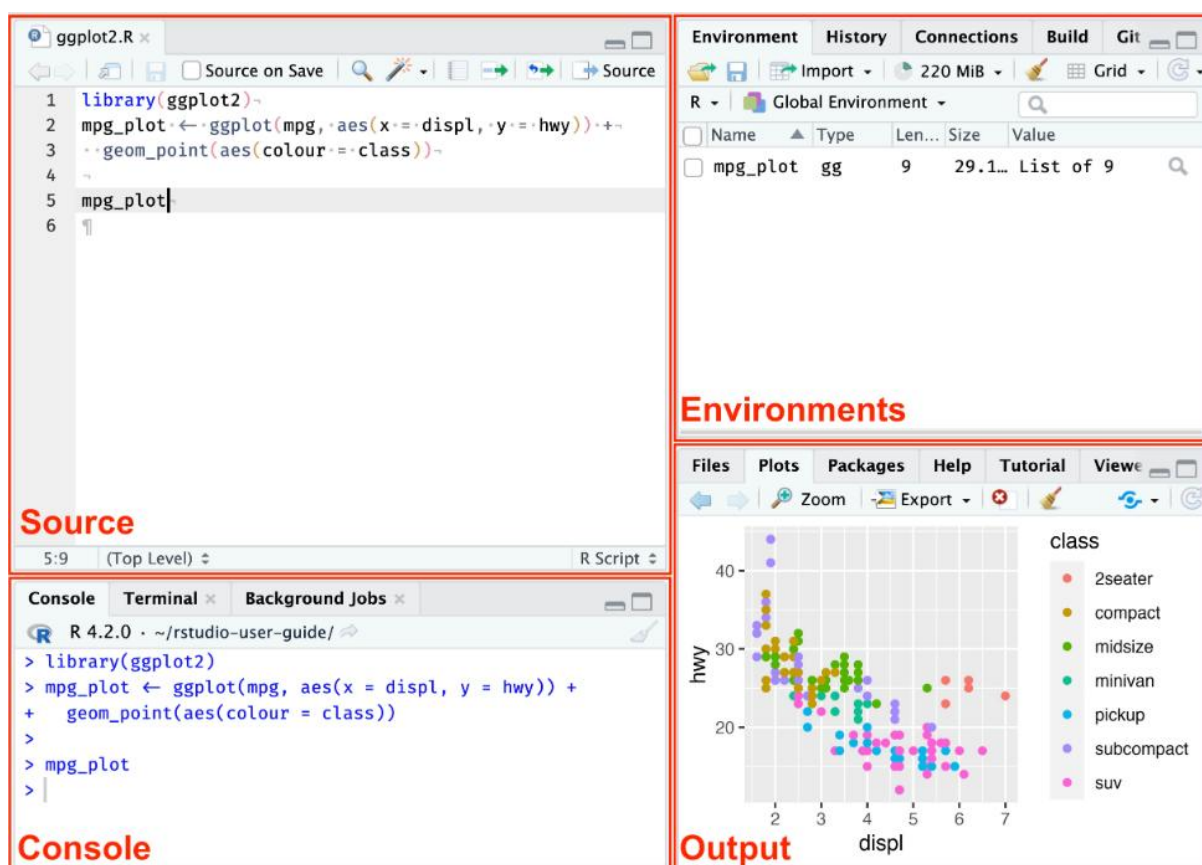


Slika 8.2 Stran za prenos programske opreme R.

Tako boste prenesli programsko opremo R, postopek namestitve pa je enak kot pri drugi programski opremi. Ko je programska oprema R nameščena, je na voljo brez naprednega



integriranega razvojnega okolja (IDE), ki je uporabnikom v pomoč in podporo pri izvajanju različnih analiz. Čeprav je mogoče opraviti katero koli analizo samo z nameščenim programom R, je bolje, da ga povežete s kakšnim sodobnim IDE, kot je RStudio, ki je eno izmed najbolj priljubljenih IDE. Postopek namestitve programa RStudio je podoben postopku namestitve osnovne programske opreme R. Pojdite na spletno stran <https://posit.co/download/rstudio-desktop/>, poiščite odprtokodno licenco RStudio Desktop, jo prenesite in namestite. Po namestitvi programov R in RStudio se uporabniku prikaže naslednji zaslon uporabniškega vmesnika (slika 8.3).



Slika 8.3 Uporabniški vmesnik in R & Rstudio (RStudio, 2024).

Uporabniški vmesnik programa RStudio ima štiri osnovne plošče (RStudio, 2024):

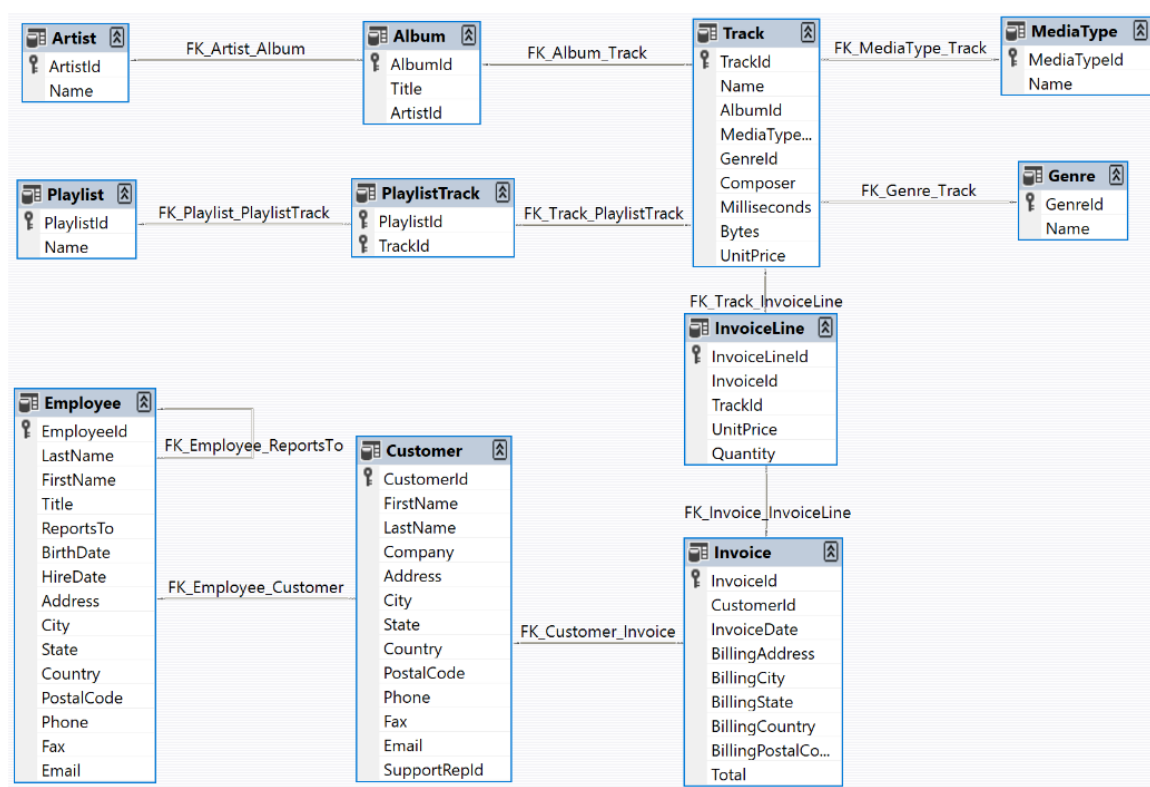
- Podokno z viri (Source);
- Podokno konzole (Console);
- Podokno Okolje, ki vsebuje zavihke Okolje, Zgodovina, Povezave, Sestavljanje, VCS in Učbenik (Environments);
- Podokno Izhod, ki vsebuje zavihke Datoteke, Sklopi, Paketi, Pomoč, Pregledovalnik in Predstavitev (Output).



Vsako od podoken ter pododdelek in možnosti v njem uporabnikom omogočajo izvajanje različnih operacij, nadzor nad nekaterimi analizami podatkov ali bolj strukturiran in jasen pregled nad procesom analize podatkov, ki se izvaja.

8.3 Kaj je SQL in kako je povezan z BA in R?

Podatkovna zbirka Chinook je vzorčna podatkovna zbirka, ki se uporablja za učenje in prikaz sistemov za upravljanje podatkovnih zbirk (DBMS) in poizvedb SQL. Zasnovana je kot trgovina z digitalnimi mediji, ki uporablja resnične podatke iz knjižnice iTunes, izmišljena imena/naslove za stranke in zaposlene ter naključne podatke za prodajne informacije. Podatkovna baza vsebuje različne tabele, ki predstavljajo podatke glasbene trgovine, vključno s podatki o izvajalcih, albumih, skladbah, strankah, računih in drugih (slika 8.4).



Slika 8.4 Podatkovni model podatkovne zbirke Chinook.

Slika 8.4 prikazuje podatkovni model podatkovne zbirke Chinook z različnimi podatkovnimi tabelami in njihovimi ključi (edinstvenimi identifikatorji) ter skupnimi tabelami, kot je tabela PlaylistTrack. Tabele posredujejo različne informacije o dani digitalni trgovini (preglednica 8.1).

**Preglednica 8.1 Informacije v vseh tabelah podatkovne zbirke Chinook.**

Table Name	Description
Artist	Contains information about music artists.
Album	Contains information about music albums, each associated with an artist.
Track	Contains information about individual music tracks, including references to albums, media types, and genres.
Genre	Contains information about music genres.
MediaType	Contains information about different types of media (e.g., audio, video).
Customer	Contains information about customers, including contact details and support representative information.
Employee	Contains information about employees, including their roles, reports-to relationships, and contact details.
Invoice	Contains information about invoices, including customer details, billing information, and total amounts.
InvoiceLine	Contains detailed information about each item in an invoice, including references to tracks and quantities.
Playlist	Contains information about playlists.
PlaylistTrack	Links tracks to playlists, indicating which tracks are included in which playlists.

8.4 Kako so povezani poslovna analitika, SQL in R?

Povezava med BA, SQL in R je naravna, saj morajo biti vsi poslovni podatki shranjeni v podatkovnih zbirkah SQL. To je še vedno idealističen cilj, saj je upravljanje podatkov v nekaterih delih malih in srednje velikih podjetij, ki še vedno ne razumejo v celoti moči podatkov, še vedno slabo. V velikih podjetjih so to že zdavnaj spoznali in podatki so ustrezno strukturirani v podatkovnih zbirkah (SQL ali drugih, vendar običajno v SQL). Po drugi strani je analizo podatkov mogoče opraviti v SQL, vendar je v ta namen bolje uporabiti statistično usmerjeno programsko opremo, kjer je v ospredju R kot ena najbolj priljubljenih statističnih platform za izvajanje analize podatkov.

Zato lahko SQL in R obravnavamo kot odlično orodje za sodelovanje, kadar gre za problem s področja BA. Glavnih razlogov je več, eden od njih pa je, da se podatki BA dnevno spreminjajo in posodablajo glede na tržno realnost in dejavnosti podjetja: prodaja, zaposleni, prihodki itd. Podatkovne zbirke SQL so odlične za zajemanje teh sprememb in posodabljanje obstoječih podatkov, medtem ko so skripte R zelo dobre za avtomatizacijo nalog in oblikovanje novih paketov za analizo danih podatkov. Razlog za to je, da je jedro R bolj zgrajeno na konceptu analiziranja podatkov kot na splošnem programiranju, kot je na primer Python.

Poizvedovanje po zbirki podatkov SQL z R



Programski jezik R in podatkovne zbirke SQL sta naravno povezana, saj je programski jezik R namenjen predvsem statistični analizi podatkov, večina podatkov o transakcijah pa je v podatkovnih zbirkah. Način, kako R upravlja manipulacijo podatkov iz podatkovnih zbirk SQL, je prek programskih paketov DBI in RSQLite. Paket DBI zagotavlja standardiziran vmesnik za interakcijo z različnimi DBMS, ki uporabnikom omogoča povezovanje, poizvedovanje in dosledno upravljanje transakcij v različnih podatkovnih zbirkah. Paket RSQLite, ki se drži vmesnika DBI, posebej olajšuje interakcijo s podatkovnimi zbirkami SQLite in uporabnikom omogoča izvajanje poizvedb SQL, pridobivanje podatkov in izvajanje operacij s podatkovnimi zbirkami neposredno iz R. Ti paketi skupaj poenostavljajo postopek dela s podatkovnimi zbirkami v R ter ponujajo skladen in učinkovit potek dela.

Da bi učinkovito prikazali izvajanje operacije SQL iz programa R in ustvarjanje želenih vpogledov v podatke v zvezi z obravnavanim problemom, smo na slikah 8.5 in 8.6 prikazali več izsekov kode.

```
---  
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)
```\n  
## Connect to the Chinook SQLite database  
  
```{r}  
con <- dbConnect(RSQLite::SQLite(), dbname = "Chinook_Sqlite.sqlite")
```\n  
## List all tables in the database  
  
```{r}  
tables <- dbListTables(con)
print(tables)
```
```

Slika 8.5 Korak kode za vzpostavitev povezave med SQL in R ter raziskovanje podatkovnih tabel, ki jih vsebuje SQL.

Prvi korak pri poizvedovanju po zbirki podatkov SQL prek programa R je vzpostavitev povezave (slika 8.5). Slika prikazuje uporabo paketov DBI in RSQLite, ki omogočata vzpostavitev povezave prek funkcije `dbConect()`. Rezultat povezave in podatkovne tabele, ki se odkrijejo prek omenjene



povezave, se nato izvozijo prek funkcij `dbListTables()`, ki izpišejo seznam vseh podatkov, najdenih prek povezave: Album, Izvajalec, Stranka, Zaposleni, Žanr, Račun, Linija računa MediaType, Seznam predvajanja, PlaylistTrack, Skladba.

Ko je povezava vzpostavljena, je mogoče izvesti številne analize, ki so odvisne od poslovnega cilja in prihodnje uporabe danih rezultatov. Na tem mestu bomo zaradi omejenega prostora prikazali le del možnih analiz podatkov z majhnim delom kode in naborom kodnih pravil, ki so potrebna za pridobivanje informacij iz SQL. Koda izvede poizvedbo po zbirki podatkov prek programa R in prikaže najbolj prodajane albume, njihove avtorje in število prodanih albumov (slika 8.6). Tabela 8.2 predstavlja rezultate poizvedovanja po podatkih prek odlomka kode na sliki 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 8.6 Delček kode za poizvedbo po SQL prek R in določitev 10 najbolj prodajanih albumov.

**Preglednica 8.2 Deset najboljše prodajanih albumov v digitalni trgovini Chinook.**

| Album title | Artist name | Quantity sold |
|--|------------------------------|---------------|
| 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 |
| My Generation - The Very Best Of The Who | The Who | 19 |
| International Superhits | Green Day | 18 |
| Chronicle, Vol. 1 | Creedence Clearwater Revival | 18 |

Reference Poglavje 8

- 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.
- Introductory Statistics. Bentham Science Publishers, Kahl, A. (Publish 2023). DOI:10.2174/97898151231351230101