

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

Laura Štefanac

Analiza onkoloških bolesnika u UAE

**PROJEKT IZ KOLEGIJA SKLADIŠTA PODATAKA I POSLOVNA
INTELIGENCIJA**

Varaždin, 2025.

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

Laura Štefanac

Studij: Baze podataka i baze znanja

Analiza onkoloških bolesnika u UAE
PROJEKT IZ KOLEGIJA SKLADIŠTA PODATAKA I POSLOVNA INTELIGENCIJA

Mentori:

Prof. dr. sc. Kornelije Rabuzin

Mag. edu. inf. Maja Cerjan

Varaždin, lipanj 2025.

Sažetak

Tema ovog rada je kreiranje skladišta podataka o onkološkim bolesnicima u Ujedinjenim Arapskim Emiratima. Korištenjem alata MySQL Workbench i Power BI prevedeni su svi koraci ETL procesa od ekstrakcije i transformacije do učitavanja podataka u bazu podataka i izrade dimenzijskog modela. U alatu Power BI su napravljene analize i vizualizacije podataka koje omogućuju bolje razumijevanje demografskih i kliničkih obilježja pacijenata te ishode liječenja.

Ključne riječi: onkologija, ETL, Power BI, MySQL Workbench, Kaggle, skladište podataka, poslovna inteligencija

Sadržaj

1. Uvod	1
2. Metode i tehnike rada.....	2
3. Izrada skladišta podataka.....	3
3.1. ETL PROCES.....	4
4. Power BI analiza	12
4.1. Izrada izvještaja.....	15
4.1.1. Izvještaj o broju slučajeva po dobnim skupinama i vrsti raka	16
4.1.2. Izvještaj o ishodima liječenja po bolnicama.....	17
4.1.3. Izvještaj o lokaciji bolnica.....	18
4.1.4. Izvještaj o ishodu tretmana pacijenata	18
4.1.5. Izvještaj o uzroku smrti	19
4.1.6. Izvještaj o pušačima/ne pušačima i ishodu liječenja.....	20
5. Zaključak.....	21
Popis literature	22
Popis slika.....	23

1. Uvod

Analiza podataka u zdravstvu danas ima ključnu ulogu u poboljšanju razumijevanja bolesti, optimizaciji terapijskih pristupa i donošenju odluka u medicinskoj praksi. S obzirom na sve veću količinu dostupnih podataka, osobito u području onkologije nužno je osigurati kvalitetnu organizaciju, integraciju i analizu. Ovaj projekt bavi se izradom skladišta podataka i analizom skupa podataka onkoloških bolesnika u Ujedinjenim Arapskim Emiratima, s ciljem prikaza procesa od ekstrakcije i transformacije do vizualizacije informacija koje su ključne pomoću suvremenih alata (MySQL Workbench, Power BI). Kroz praktičnu primjenu ETL procesa i izradu interaktivnih izvještaja, projekt pokazuje dublji uvid u obrasce liječenja i ishode kod različitih vrsta karcinoma.

Poveznica na videozapis: https://youtu.be/rTWd8wbzypY?si=rz99hlqllp7lF_-e

2. Metode i tehnike rada

U nastavku će biti opisane korištene metode i tehnike rada za ovaj projektni zadatak. Skup podataka preuzet je s platforme Kaggle pod nazivom UAE Cancer Patient Dataset. (Kaggle, n.d.) Skup podataka je pohranjen u .csv formatu datoteke, sadrži podatke o pacijentima oboljelima od raka u Ujedinjenim Arapskim Emiratima, uključuje demografske podatke, podatke o dijagnozama, tretmanima, bolnicama, liječnicima, statusu pušenja i ishodima liječenja.

Za obradu podataka korišteni su: Microsoft Excel, MySQL Workbench i Power BI. Microsoft Excel je korišten u početnoj fazi kako bi se izvršila provjera podataka. Za ETL proces je korišten MySQL Workbench, alat za rad s MySQL bazama podataka koji omogućuje dizajniranje, modeliranje, administraciju i upite nad bazom podataka u grafičkom sučelju. (MySQL).

Za kreiranje izvještaja korišten je Power BI, alat za vizualizaciju podataka koji omogućuje povezivanje s različitim izvorima podataka, njihovu obradu te izradu interaktivnih izvještaja i nadzornih ploča. Alat daje rješenja koja kombiniraju više komponenti, od dohvaćanja i transformacije podataka do njihove analize i prikaza kroz ranije interaktivne oblike poput grafova, karata i tabličnih prikaza. (Microsoft, n.d.)

3. Izrada skladišta podataka

Zdravstvene ustanove i istraživački centri sve više se oslanjaju na podatke kako bi unaprijedili razumijevanje u bolesti, optimizirali terapijske pristupe i poboljšali ishode liječenja. Velike količine podataka se prikupljaju iz različitih izvora npr. elektronički zdravstveni zapisi, laboratorijski nalaza, registara pacijenata te sustava za upravljanje bolnicama. Takvi podaci mogu biti strukturirani ili nestrukturirani, uključuju informacije o dijagnozama, terapijama, demografskim obilježjima pacijenata, navikama i životnom stilu, posebno su korisni u prepoznavanju obrazaca i donošenju medicinskih i organizacijskih odluka temeljenih na dokazima.

Ključno je pravilno oblikovati skladište podataka koje omogućuje kvalitetnu organizaciju i kasniju analizu prikupljenih informacija. U ovom projektu korišten je skup podataka *UAE Cancer Patient Dataset* koji sadrži podatke o pacijentima oboljelima od raka u Ujedinjenim Arapskim Emiratima. (Kaggle, n.d.) Ovaj skup podataka uključuje širok spektar atributa, kako bi se omogućila učinkovita analiza i vizualizacija ovih podataka, bilo je potrebno osmisliti strukturirano skladište koje omogućuje njihovu integraciju, konzistentnost i dostupnost za daljnju obradu. Skup podataka sastoji se od 10 000 zapisa o pacijentima. Podaci koji dolaze u .csv datoteci su sljedeći:

- *Patient_ID* – jedinstveni identifikator svakog pacijenta
- *Age* – dob pacijenta
- *Gender* – spol pacijenta
- *Nationality* – nacionalnost pacijenta
- *Emirate* – teritorijalna jedinica u UAE u kojoj je pacijent liječen
- *Diagnosis_Date* – datum postavljanja dijagnoze
- *Cancer_Type* – vrsta karcinoma
- *Cancer_Stage* – stadij karcinoma (I, II, III, IV)
- *Treatment_Type* – vrsta tretmana primijenjena na pacijentu (kemoterapija, radioterapija, kirurgija)
- *Treatment_Start_Date* – datum početka liječenja
- *Hospital* – naziv bolnice u kojoj je pacijent liječen
- *Primary_Physician* – glavni liječnik zadužen za liječenje pacijenta
- *Outcome* – ishod liječenja (izliječen, u tijeku, preminuo)
- *Cause_of_Death* – uzrok smrti
- *Smoking_Status* – navike pušenja (pušač, nepušač, bivši pušač)

- *Comorbidities* – opis komorbiditeta, odnosno drugih zdravstvenih stanja uz rak
- *Ethnicity* – etnička pripadnost pacijenta
- *Weight* – tjelesna masa pacijenta
- *Height* – visina pacijenta

Slika 1: Prikaz podataka u Excel-u, snimka zaslona(2025.)

Nedostajući podaci postoje nedostajući podaci u atributima *Death_Date*(9008), *Cause_of_Death*(9008) i *Comorbidities*(4047), nedostajuće vrijednosti vidljive su na Kaggle stranici odakle je preuzet skup podataka. Na slici 1 vidi se prikaz kako skup podataka izgleda nakon preuzimanja u Excelu, podaci su teško čitljivi u ovakvom formatu, stoga je potreban ETL proces za njihovu obradu i analizu.

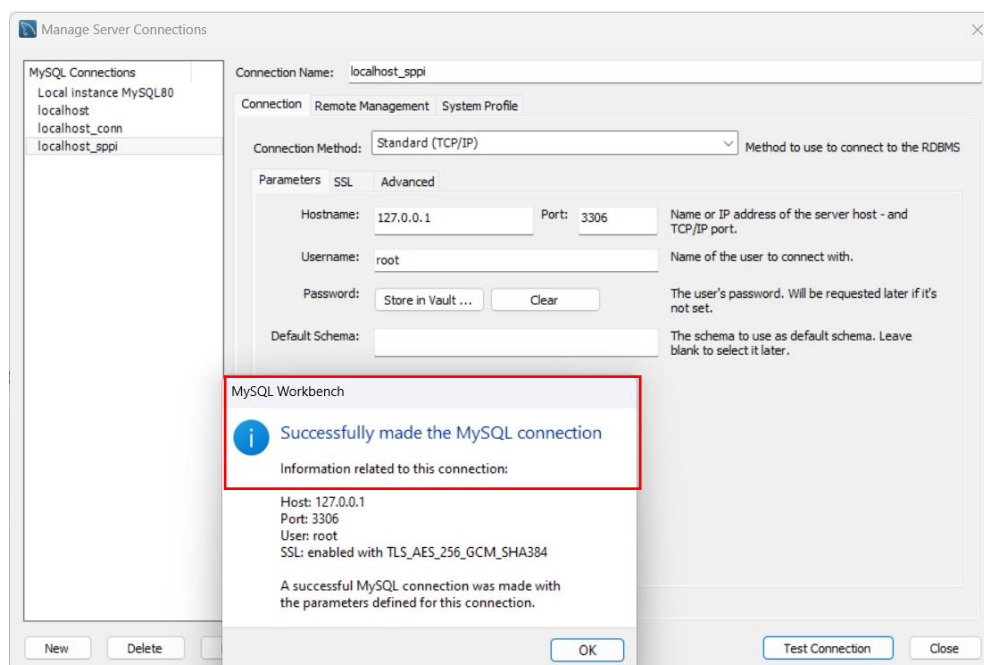
3.1. ETL PROCES

ETL, odnosno ekstrakcija, transformacija i učitavanje podataka je sljedeći korak u projektu. ETL proces je rađen u MySQL Workbench aplikaciji. Prvo je potrebno kreirati lokalni server kako bismo mogli uspješno napraviti ETL proces. Klikom na simbol „+“ pored MySQL Connections, otvara nam se novi prozor u kojem upisujemo ime konekcije. U projektu to je *localhost_sppi*, ostale podatke ostavljamo onako kako su već unesene, *hostname*, *port*, *username*. Zatim testiramo našu konekciju, i ako je sve dobro ispunjeno dobivamo potvrdnu poruku koja je vidljiva na slici 4. Sljedeći korak je kreiranje sheme pod nazivom *cancer_dw*. Nakon što je shema kreirana kreira se privremena tablica *staging_table* u koju će biti uvezeni podatke iz .csv datoteke koja je ranije preuzeta. Kada su podaci uvezeni u *staging_table*, kreiraju se tablice: *patient*, *date_dim*, *location*, *cancer* i *treatment*, nakon toga podaci se iz *staging_table* uvoze u svaku od navedenih tablica. Za nedostajuće vrijednosti u atributima *comorbidities* sve nedostajuće vrijednosti zamijenjene su tekстом 'None', dok kod atributa

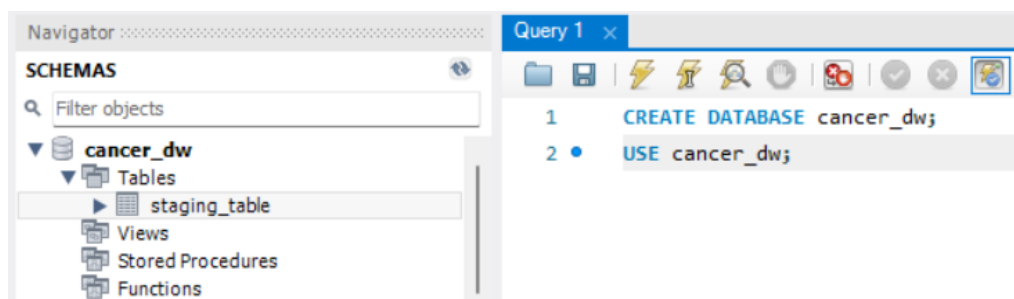
death_date je zamijenjeno s vrijednosti NULL i kod *cause_of_death* nedostajuće vrijednosti su zamijenjene su s 'N/A'.



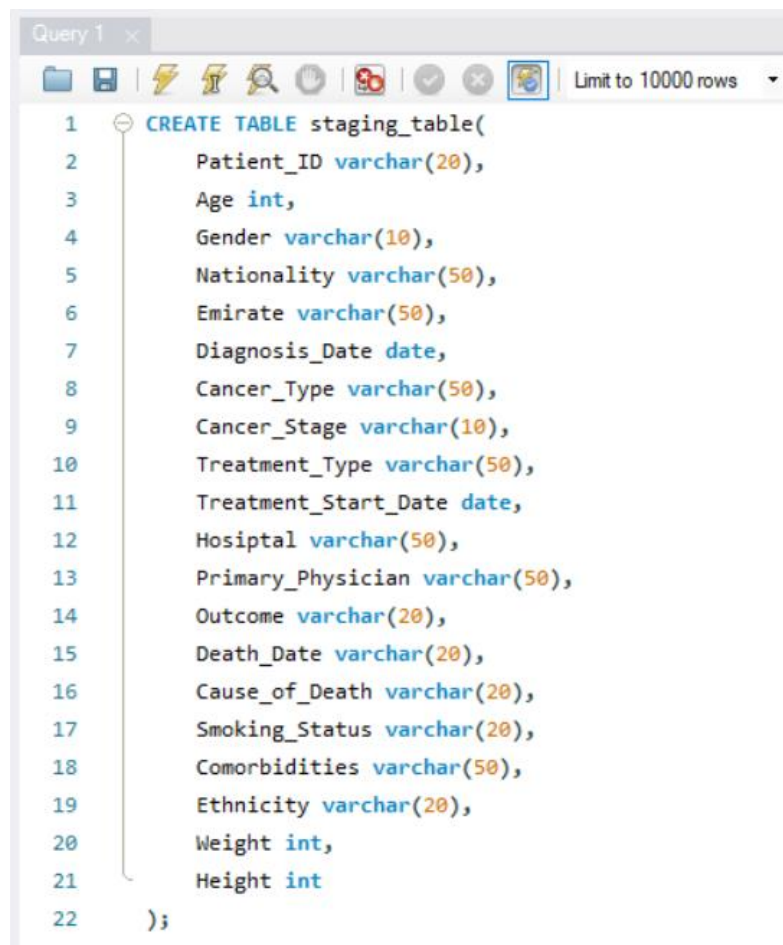
Slika 2: Kreiranje lokalnog servera, snimka zaslona(2025.)



Slika 3: Kreiranje i testiranje konekcije za lokalni poslužitelj, snimka zaslona(2025.)



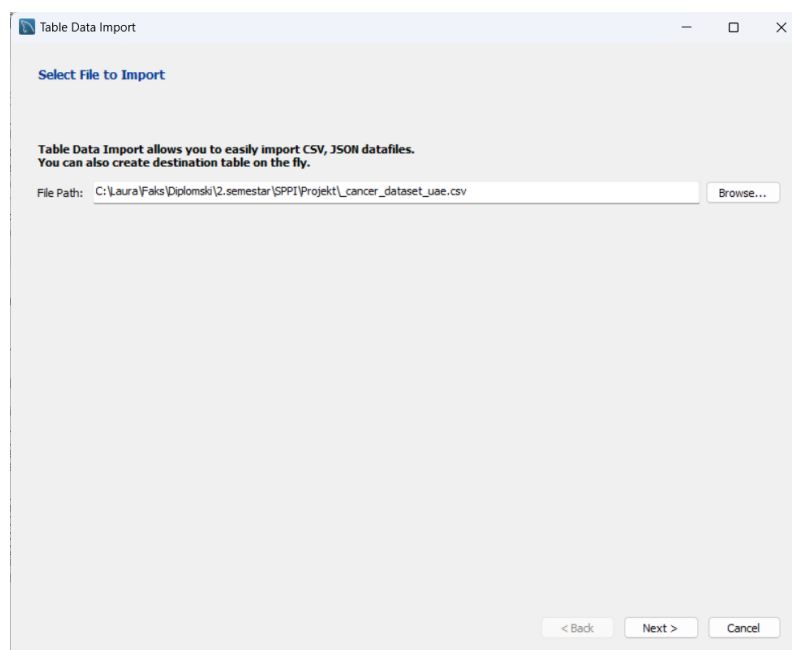
Slika 4: Kreiranje sheme cancer_dw, snimka zaslona(2025.)



The screenshot shows a SQL query editor window titled "Query 1". The query is a CREATE TABLE statement for a table named "staging_table". The table has 20 columns with the following data types: Patient_ID (varchar(20)), Age (int), Gender (varchar(10)), Nationality (varchar(50)), Emirate (varchar(50)), Diagnosis_Date (date), Cancer_Type (varchar(50)), Cancer_Stage (varchar(10)), Treatment_Type (varchar(50)), Treatment_Start_Date (date), Hosiptal (varchar(50)), Primary_Physician (varchar(50)), Outcome (varchar(20)), Death_Date (varchar(20)), Cause_of_Death (varchar(20)), Smoking_Status (varchar(20)), Comorbidities (varchar(50)), Ethnicity (varchar(20)), Weight (int), and Height (int). The query is numbered 1 through 22. The editor has a toolbar with various icons and a "Limit to 10000 rows" dropdown menu.

```
1 CREATE TABLE staging_table(  
2     Patient_ID varchar(20),  
3     Age int,  
4     Gender varchar(10),  
5     Nationality varchar(50),  
6     Emirate varchar(50),  
7     Diagnosis_Date date,  
8     Cancer_Type varchar(50),  
9     Cancer_Stage varchar(10),  
10    Treatment_Type varchar(50),  
11    Treatment_Start_Date date,  
12    Hosiptal varchar(50),  
13    Primary_Physician varchar(50),  
14    Outcome varchar(20),  
15    Death_Date varchar(20),  
16    Cause_of_Death varchar(20),  
17    Smoking_Status varchar(20),  
18    Comorbidities varchar(50),  
19    Ethnicity varchar(20),  
20    Weight int,  
21    Height int  
22 );
```

Slika 5: Kreiranje privremene tablice za uvoz podataka, snimka zaslona(2025.)



Slika 6: Uvoz podataka pomoću čarobnjaka, snimka zaslona(2025.)

Table Data Import

Import Data

The following tasks will now be performed. Please monitor the execution.

☒ Prepare Import
 ☒ Import data file

Finished performing tasks. Click [Next >] to continue.

Show Logs

< Back

Next >

Cancel

Slika 7: Uvoz podataka u tablicu, snimka zaslona(2025.)

1 • SELECT * FROM cancer_dw.staging_table;

Result Grid

Filter Rows:

Export:

Wrap Cell Contents:

Fetch rows:

Patient_ID	Age	Gender	Nationality	Emirate	Diagnosis_Date	Cancer_Type	Cancer_Stage	Treatment_Type	Treatment_Start_Date	Hospital	Primary_Physician	Outcome
PAT000001	69	Female	Emirati	Umm Al Quwain	2020-11-30	Liver	II	Radiation	2020-12-04	Sheikh Khalifa Hospital	Dr. VO41	Recovered
PAT000002	32	Male	Emirati	Umm Al Quwain	2015-10-10	Leukemia	III	Surgery	2015-11-05	Dubai Hospital	Dr. SM31	Recovered
PAT000003	89	Male	Emirati	Abu Dhabi	2018-02-13	Liver	III	Radiation	2018-08-03	Zayed Military Hospital	Dr. BC7	Under Treatment
PAT000004	78	Female	Emirati	Abu Dhabi	2022-02-04	Lung	III	Radiation	2022-03-13	Cleveland Clinic Abu Dhabi	Dr. TC14	Recovered
PAT000005	38	Female	Emirati	Fujairah	2019-12-03	Pancreatic	II	Chemotherapy	2020-02-29	Sheikh Khalifa Hospital	Dr. YS37	Recovered
PAT000006	41	Female	Emirati	Abu Dhabi	2015-02-10	Breast	I	Surgery	2015-04-21	Dubai Hospital	Dr. ZL30	Recovered
PAT000007	20	Male	Emirati	Ras Al Khaimah	2023-02-27	Ovarian	IV	Chemotherapy	2023-05-15	Sheikh Khalifa Hospital	Dr. AL45	Recovered
PAT000008	39	Male	Expatriate	Sharjah	2018-04-05	Breast	I	Surgery	2018-08-10	Sheikh Khalifa Hospital	Dr. IL20	Under Treatment
PAT000009	70	Female	Emirati	Umm Al Quwain	2022-12-12	Lung	III	Immunotherapy	2023-05-29	Dubai Hospital	Dr. KH33	Deceased
PAT000010	19	Female	Emirati	Dubai	2018-05-29	Breast	I	Surgery	2018-07-25	Cleveland Clinic Abu Dhabi	Dr. GX45	Recovered
PAT000011	47	Male	Emirati	Abu Dhabi	2023-02-21	Prostate	II	Chemotherapy	2023-07-27	Dubai Hospital	Dr. IH48	Under Treatment
PAT000012	55	Female	Expatriate	Umm Al Quwain	2015-11-09	Colorectal	III	Radiation	2016-05-05	Zayed Military Hospital	Dr. NS1	Recovered
PAT000013	19	Female	Expatriate	Ras Al Khaimah	2020-04-14	Lung	II	Surgery	2020-07-18	Zayed Military Hospital	Dr. FY12	Under Treatment
PAT000014	81	Male	Expatriate	Ras Al Khaimah	2021-11-23	Liver	IV	Radiation	2022-01-24	Cleveland Clinic Abu Dhabi	Dr. FM46	Under Treatment
PAT000015	77	Female	Emirati	Fujairah	2018-02-28	Ovarian	III	Surgery	2018-03-18	Sheikh Khalifa Hospital	Dr. LT23	Under Treatment
PAT000016	38	Female	Emirati	Abu Dhabi	2015-02-17	Colorectal	IV	Chemotherapy	2015-07-21	Cleveland Clinic Abu Dhabi	Dr. WV1	Recovered
PAT000017	50	Female	Expatriate	Umm Al Quwain	2015-05-14	Breast	III	Immunotherapy	2015-08-15	Cleveland Clinic Abu Dhabi	Dr. QX39	Recovered
PAT000018	75	Female	Emirati	Umm Al Quwain	2022-01-10	Ovarian	I	Immunotherapy	2022-05-03	Sheikh Khalifa Hospital	Dr. WC13	Deceased
PAT000019	39	Female	Emirati	Fujairah	2019-04-29	Leukemia	II	Immunotherapy	2019-06-20	Dubai Hospital	Dr. QD15	Recovered
PAT000020	66	Male	Expatriate	Umm Al Quwain	2023-05-05	Pancreatic	III	Radiation	2023-08-09	Sheikh Khalifa Hospital	Dr. JX8	Recovered
PAT000021	76	Male	Emirati	Ajman	2021-06-22	Colorectal	I	Chemotherapy	2021-12-11	Cleveland Clinic Abu Dhabi	Dr. LU7	Recovered
PAT000022	59	Male	Emirati	Ajman	2020-11-03	Pancreatic	IV	Surgery	2020-11-13	Zayed Military Hospital	Dr. UZ28	Under Treatment
PAT000023	77	Male	Emirati	Ajman	2017-12-07	Pancreatic	III	Chemotherapy	2018-04-05	Dubai Hospital	Dr. FJ33	Under Treatment
PAT000024	32	Male	Emirati	Abu Dhabi	2023-05-16	Leukemia	III	Surgery	2023-08-28	Dubai Hospital	Dr. PY41	Under Treatment

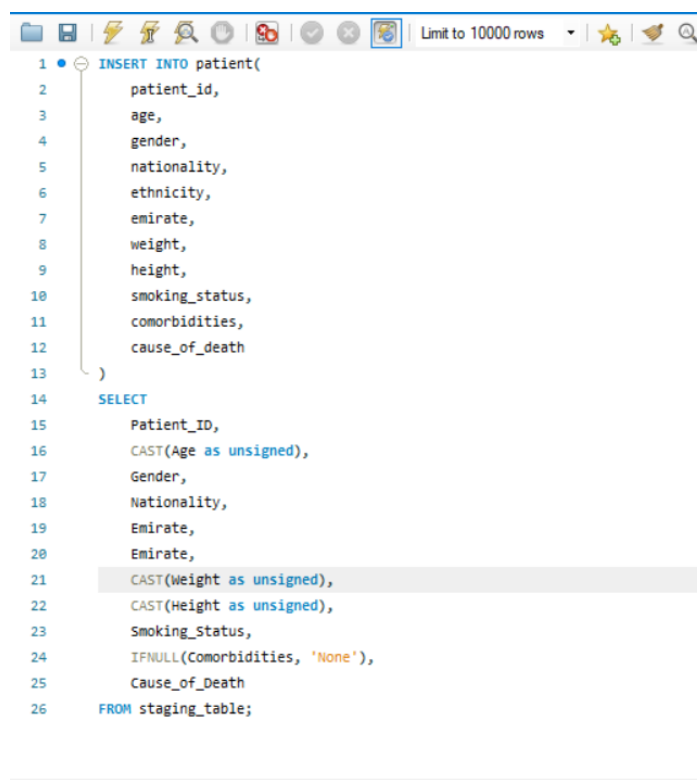
Slika 8: Prikaz uvezenih podataka, snimka zaslona(2025.)

```

1  ● ○ CREATE TABLE patient (
2      patient_id VARCHAR(20) PRIMARY KEY,
3      age INT,
4      gender VARCHAR(10),
5      nationality VARCHAR(50),
6      ethnicity VARCHAR(50),
7      emirate VARCHAR(50),
8      weight INT,
9      height INT,
10     smoking_status VARCHAR(30),
11     comorbidities VARCHAR(100),
12     cause_of_death VARCHAR(100)
13 );

```

Slika 9: Kreiranje tablice patient, snimka zaslona(2025.)



```

1  ● ○ INSERT INTO patient(
2      patient_id,
3      age,
4      gender,
5      nationality,
6      ethnicity,
7      emirate,
8      weight,
9      height,
10     smoking_status,
11     comorbidities,
12     cause_of_death
13 )
14 SELECT
15     Patient_ID,
16     CAST(Age as unsigned),
17     Gender,
18     Nationality,
19     Emirate,
20     Emirate,
21     CAST(Weight as unsigned),
22     CAST(Height as unsigned),
23     Smoking_Status,
24     IFNULL(Comorbidities, 'None'),
25     Cause_of_Death
26 FROM staging_table;

```

Slika 10: Unos podataka u tablicu patient iz staging_table, snimka zaslona(2025.)

```

1 CREATE TABLE date_dim (
2     date_id INT AUTO_INCREMENT PRIMARY KEY,
3     full_date DATE,
4     year INT,
5     month INT,
6     day INT,
7     weekday VARCHAR(10)
8 );

```

Slika 11: Kreiranje tablice date_dim, snimka zaslona(2025.)

```

1 CREATE TABLE cancer (
2     cancer_id INT AUTO_INCREMENT PRIMARY KEY,
3     cancer_type VARCHAR(50),
4     cancer_stage VARCHAR(10)
5 );
6

```

Slika 12: Kreiranje tablice cancer, snimka zaslona(2025.)

```

CREATE TABLE location (
    location_id INT AUTO_INCREMENT PRIMARY KEY,
    hospital VARCHAR(100),
    primary_physician VARCHAR(50),
    emirate VARCHAR(50)
);

```

Slika 13: Kreiranje tablice location, snimka zaslona(2025.)

```

1 CREATE TABLE treatment (
2     treatment_id INT AUTO_INCREMENT PRIMARY KEY,
3     diagnosis_date INT,
4     treatment_date INT,
5     death_id INT,
6     patient_id VARCHAR(20),
7     cancer_id INT,
8     location_id INT,
9     treatment_type VARCHAR(50),
10    outcome VARCHAR(50)
11 );

```

Slika 14: Kreiranje tablice treatment, snimka zaslona(2025.)

```

1 • INSERT INTO cancer (cancer_type, cancer_stage)
2   SELECT DISTINCT
3     Cancer_Type,
4     Cancer_Stage
5   FROM staging_table;

```

Slika 15: Unos podataka u tablicu cancer iz staging_table, snimka zaslona(2025.)

```

1 • INSERT INTO date_dim (full_date, year, month, day, weekday)
2   SELECT DISTINCT
3     CAST(date_str AS DATE) AS full_date,
4     YEAR(CAST(date_str as date)),
5     MONTH(CAST(date_str as date)),
6     DAY(CAST(date_str as date)),
7     DAYNAME(CAST(date_str as date))
8   FROM
9   (
10    SELECT Diagnosis_Date AS date_str FROM staging_table
11    UNION
12    SELECT Treatment_Start_Date AS date_str FROM staging_table
13    UNION
14    SELECT Death_Date AS date_str FROM staging_table WHERE Death_Date IS NOT NULL AND Death_Date != 'N/A'
15  ) AS all_dates
16  WHERE date_str IS NOT NULL AND date_str != 'N/A';

```

Slika 16: Unos podataka u tablicu dim_date iz staging_table, snimka zaslona(2025.)

```

• INSERT INTO location (hospital, primary_physician, emirate)
  SELECT DISTINCT
    Hospital,
    Primary_Physician,
    Emirate
  FROM staging_table;

```

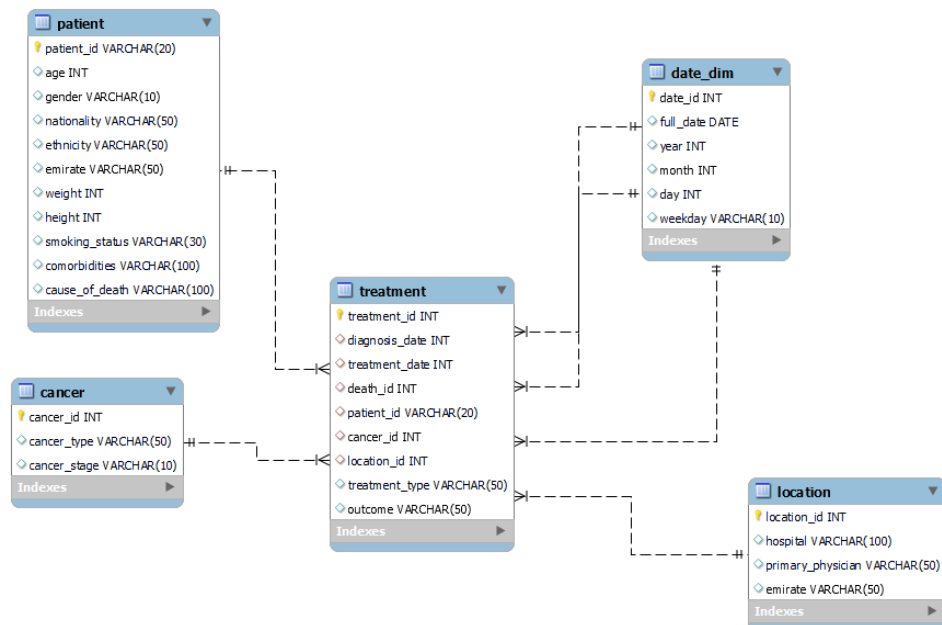
Slika 17: Unos podataka u tablicu location iz staging_table, snimka zaslona(2025.)

```

1  INSERT INTO treatment (
2      diagnosis_date,
3      treatment_date,
4      death_id,
5      patient_id,
6      cancer_id,
7      location_id,
8      treatment_type,
9      outcome
10 )
11 SELECT
12     d1.date_id AS diagnosis_date,
13     d2.date_id AS treatment_date,
14     d3.date_id AS death_date,
15     s.Patient_ID,
16     c.cancer_id,
17     l.location_id,
18     s.Treatment_Type,
19     s.Outcome
20 FROM staging_table s
21
22 JOIN date_dim d1 ON d1.full_date = CAST(s.Diagnosis_Date AS date)
23 JOIN date_dim d2 ON d2.full_date = CAST(s.Treatment_Start_Date AS date)
24 LEFT JOIN date_dim d3
25     ON d3.full_date =
26     CASE
27         WHEN s.Death_Date IS NULL OR s.Death_Date = 'N/A' THEN NULL
28         ELSE CAST(s.Death_Date AS date)
29     END
30 JOIN cancer c ON c.cancer_type = s.Cancer_Type AND c.cancer_stage = s.Cancer_Stage
31 JOIN location l ON l.hospital = s.Hospital AND l.primary_physician = s.Primary_Physician;

```

Slika 18: Unos podataka u tablicu treatment iz staging_table, snimka zaslona(2025.)

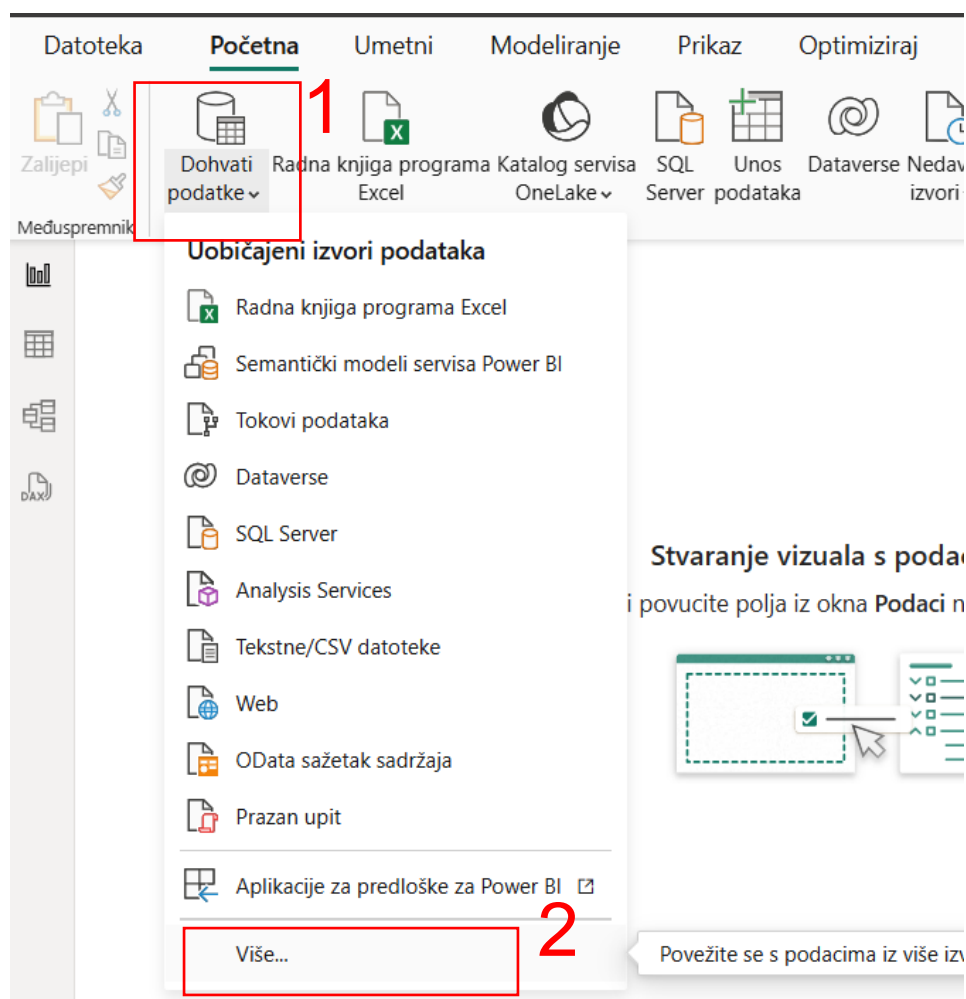


Slika 19: ERA model, snimka zaslona (2025.)

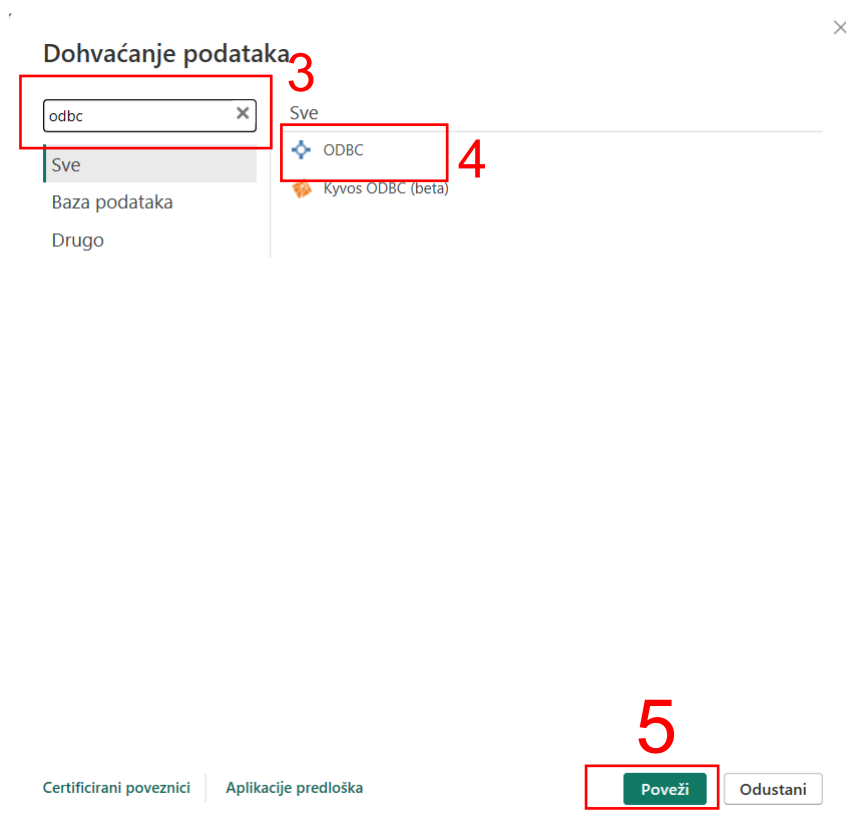
4. Power BI analiza

Power BI je Microsoftova platforma za poslovnu inteligenciju koja omogućuje povezivanje s različitim izvorima podataka, njihovu transformaciju te izradu interaktivnih vizualizacija i izvještaja. Alat je namijenjen korisnicima svih razina znanja i koristi se za brzo otkrivanje i prikazivanje ključnih poslovnih uvida kroz pregledne grafove, mape i *dashboards*.(Microsoft, 2025.) Power BI je vrlo intuitivan i jednostavan za korištenje.

Na sljedećim slikama prikazan je postupak dohvaćanja podataka. Za dohvaćanje podataka korišten je ODBC (engl. *Open Database Connectivity*) jer Power BI Desktop ne podržava izvorno povezivanje na MySQL bazu podataka bez dodatnog konektora. ODBC je preuzet s MySQL web stranice, gdje su dostupni službeni alati za povezivanje s bazom. (Oracle, n.d.) Nakon odabira načina povezivanja, odabire se naziv izvora podataka (*cancer_dw*), a zatim i tablice koje želimo učitati.



Slika 20: Odabir dohvaćanja podataka, snimka zaslona(2025.)



Slika 21: Odabir dohvaćanja podataka (vrsta povezivanja), snimka zaslona(2025.)



Slika 22: Odabir izvora podataka, snimka zaslona(2025.)

Navigator

Prikaz mogućnosti

ODBC (dsn=cancer_odbc) [6]

- cancer_dw [6]
 - ☒ cancer
 - ☒ date_dim
 - ☒ location
 - ☒ patient
 - ☐ staging_table
 - ☒ treatment
 - information_schema
 - iznajmljivanje_automobila
 - mysql
 - performance_schema
 - sys

treatment

treatment_id	diagnosis_date	treatment_date	death_id	patient_id
2	566	1	null	PAT000637
3	1023	2	3103	PAT001248
4	681	4	null	PAT000779
5	167	7	null	PAT000170
6	23	8	null	PAT000023
7	618	10	null	PAT000698
8	790	12	null	PAT000918
9	30	12	null	PAT000358
10	877	15	null	PAT001043
11	735	15	null	PAT000847
12	425	16	null	PAT000457
13	524	18	null	PAT000584
14	53	18	null	PAT000054
15	550	19	null	PAT000618
16	125	19	null	PAT000878
17	949	22	null	PAT001141
18	572	22	null	PAT001210
19	602	23	null	PAT000678
20	352	33	null	PAT000373
21	653	35	null	PAT001373
22	77	36	2413	PAT000078
23	806	39	null	PAT000939
24	560	39	null	PAT000628

Odabir povezanih tablica

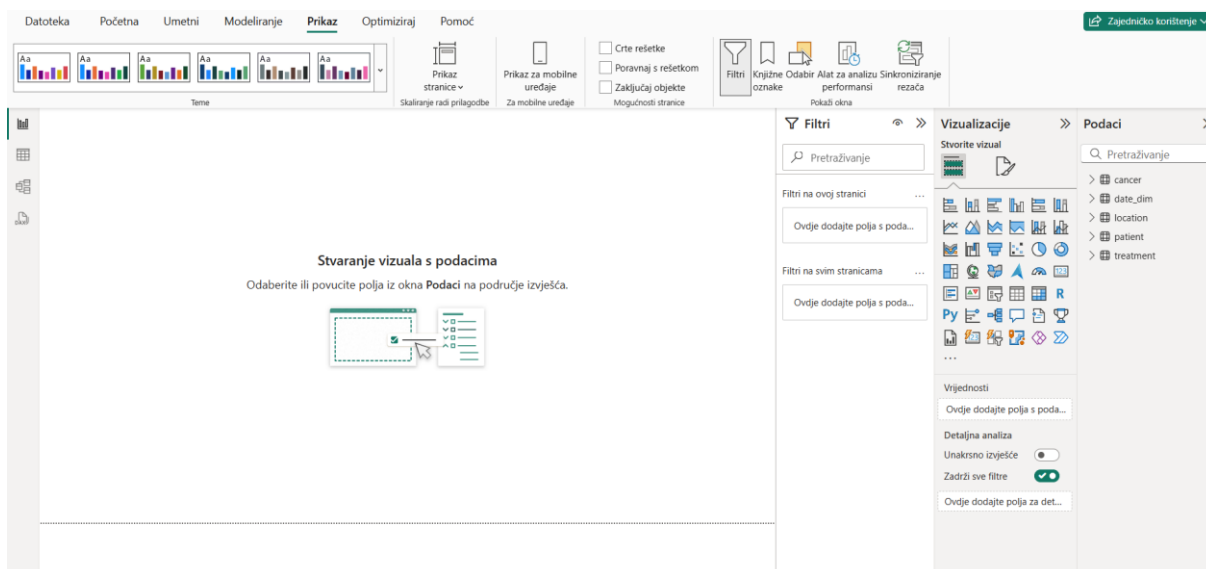
9

Učitaj

Transformacija podataka

Odustani

Slika 23: Odabir tablica za učitavanje u Power BI, snimka zaslona(2025.)



Slika 24: Početno stanje nakon učitavanja tablica, snimka zaslona(2025.)

4.1. Izrada izvještaja

Nakon što su podaci uvezeni i pripremljeni, sljedeći korak je izrada izvješća koja omogućuje detaljnu analizu i vizualizaciju ključnih informacija iz baze podataka. U sljedećim poglavljima prikazana su gotova izvješća koja su izrađena u Power BI-u.

Za dodatnu bolju analizu dodana su dva nova stupca `dobnaSkupina` i `brojPacijenataPoTretmanu`, kreirani su pomoću DAX izraza.(Microsoft, n.d.) `dobnaSkupina` dodana je u tablicu *patient*, kreirana je kako bi se lakše analizirali demografski podaci, odnosno skupine pacijenata.

`dobnaSkupina =`

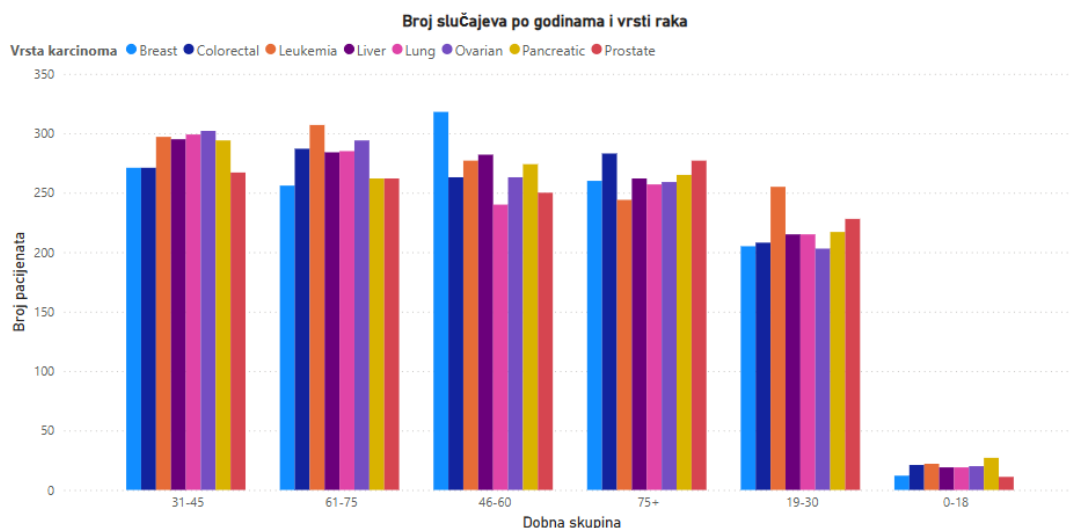
```
SWITCH (
    TRUE(),
    patient[age] <= 18, "0-18",
    patient[age] <= 30, "19-30",
    patient[age] <= 45, "31-45",
    patient[age] <= 60, "46-60",
    patient[age] <= 75, "61-75",
    "75+"
)
```

`brojPacijenataPotretmanu` je kreiran unutar tablice *treatment* kako bi se izračunao broj pacijenata za svaki tretmana.

```
brojPacijenataPoTretmanu = DISTINCTCOUNT(treatment[patient_id])
```

4.1.1. Izvještaj o broju slučajeva po dobnim skupinama i vrsti raka

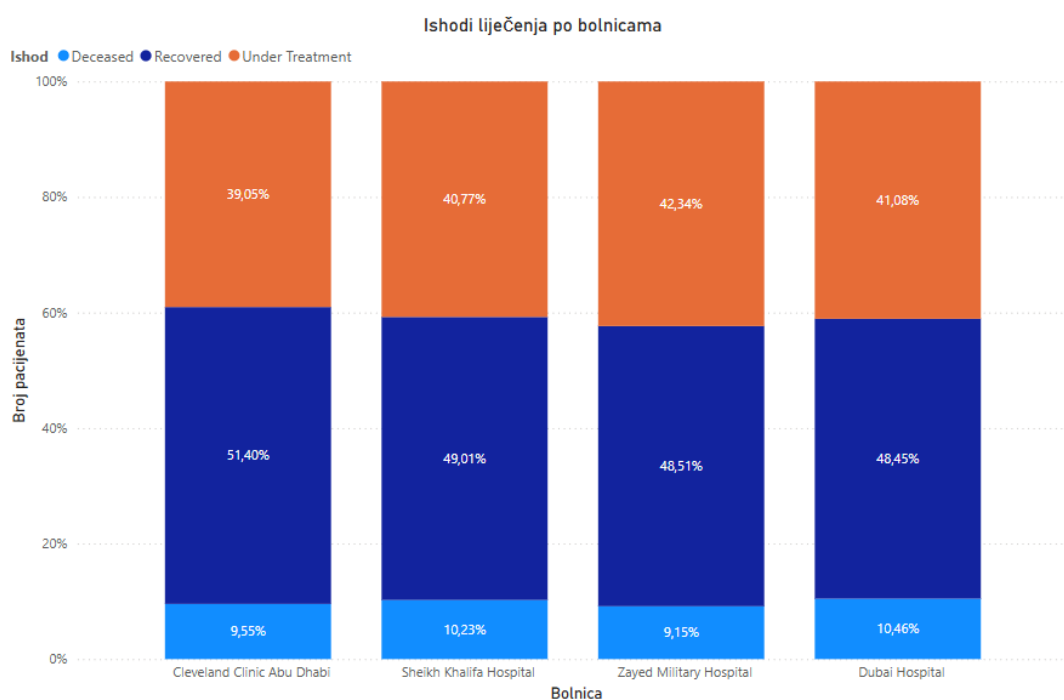
Prvi izvještaj prikazuje broj slučajeva po dobnim skupinama i vrsti raka. Najveći broj oboljelih zabilježen je u dobnj skupni 31-45, dok je najmanji broj zabilježenih slučajeva je u dobnj skupini 0-18. U dobnj skupini 31-45 najzastupljeniji je karcinom jajnika, dok je karcinom prostate najmanje zastupljen. U skupini 61-75 najviše je zastupljena leukemija, a najmanje rak dojke. Dobna skupina 46-60 ima najzastupljeniji karcinom dojke, dok najmanje je zastupljen karcinom pluća. Dobna skupina 75+ ima najzastupljeniji karcinom debelog crijeva, dok najmanje je zastupljena leukemija. U dobnj skupini 19-30 najzastupljenija je leukemija, a najmanje zastupljena je karcinom jajnika.



Slika 25:Izvještaj o broju slučajeva po dobnim skupinama i vrsti raka, snimka zaslona(2025.)

4.1.2. Izvještaj o ishodima liječenja po bolnicama

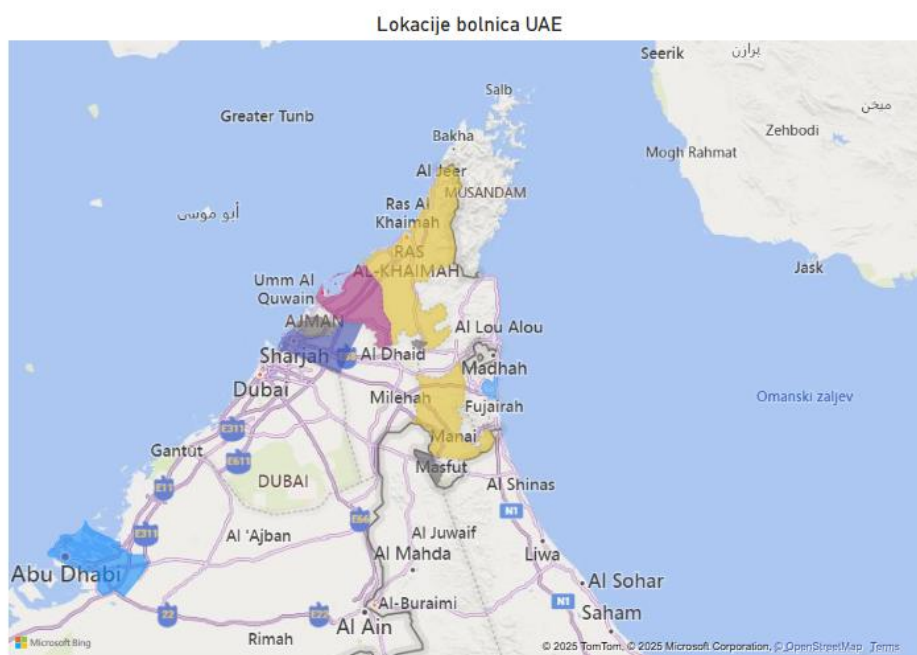
Drugi izvještaj prikazuje ishode liječenja po bolnicama, raspoređene u tri kategorije preminuli (*Deceased*), izliječeni (*Recovered*) i pacijenti na trenutnom liječenju (*Under Treatment*). Najveći udio u svim bolnicama čine izliječeni pacijenti (48-51%), dok je udio preminulih najmanji i kreće se između 9% i 10%. Udio pacijenata koji su još uvijek na liječenju iznosi između 39% i 42%. Najveći broj izliječenih pacijenata je u bolnici Cleveland Clinic Abu Dhabi iznosi 51,40%, najveći broj pacijenata koji je i dalje na liječenju je u bolnici Zayed Military Hospital i iznosi 42,34%, najveći broj preminulih pacijenata je u bolnici Dubai Hospital i iznosi 10,46%.



Slika 26: Izvještaj o ishodima liječenja po bolnicama, snimka zaslona(2025.)

4.1.3. Izvještaj o lokaciji bolnica

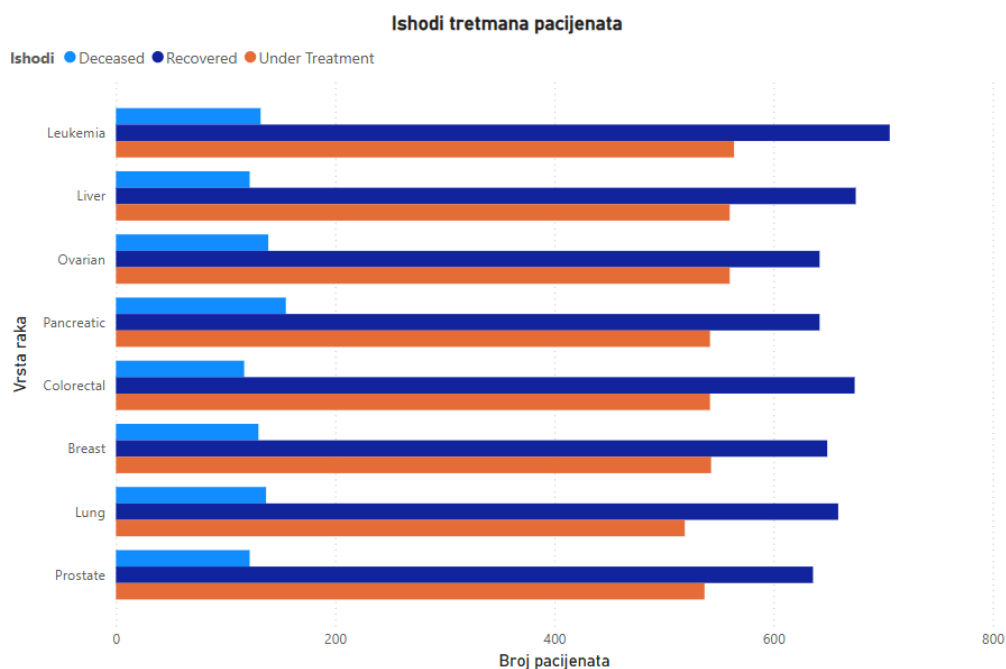
Na prikazanoj karti vide se geografske lokacije koje su uključene u analizu skupa podataka. Svaka bolnica je prikazana unutar teritorijalne jedinice (*emirate*) kojem pripada, a time se dobiva i regionalni kontekst obrade podataka. Prikazane bolnice nalaze se u sljedećim emiratima: Abu Dhabi, Ajman, Dubai, Fujairah, Ras Al Khaimah, Sharjah, Umm Al Quwain. Iz prikaza možemo uočiti prostorne distribucije bolnica i područja koje pokrivaju.



Slika 27: Izvještaj o lokacijama bolnica, snimka zaslona(2025.)

4.1.4. Izvještaj o ishodu tretmana pacijenata

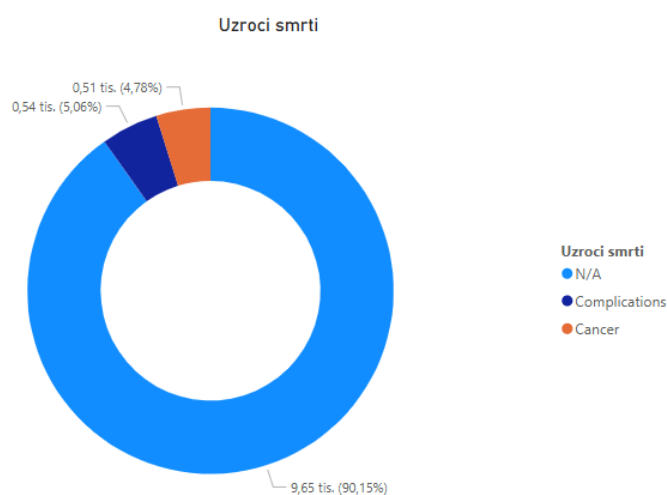
Ovaj izvještaj prikazuje ishode liječenja pacijenata za svaku pojedinu vrstu karcinoma. Ishodi su prikazani kroz tri kategorije: preminuli, oporavljeni, pacijenti koji se još uvijek liječe. Na temelju grafa zaključuje se sljedeće, najveći broj oporavljenih zabilježen je kod leukemije i karcinoma jetre, dok karcinom gušterače i pluća ima najveći udio preminulih. Većina karcinoma bilježi dobar omjer oporavka, a broj pacijenata koji su još uvijek na liječenju značajan je kod svih dijagnoza.



Slika 28: Izvještaj o ishodima tretmana pacijenata po vrsti karcinoma, snimka zaslona(2025.)

4.1.5. Izvještaj o uzroku smrti

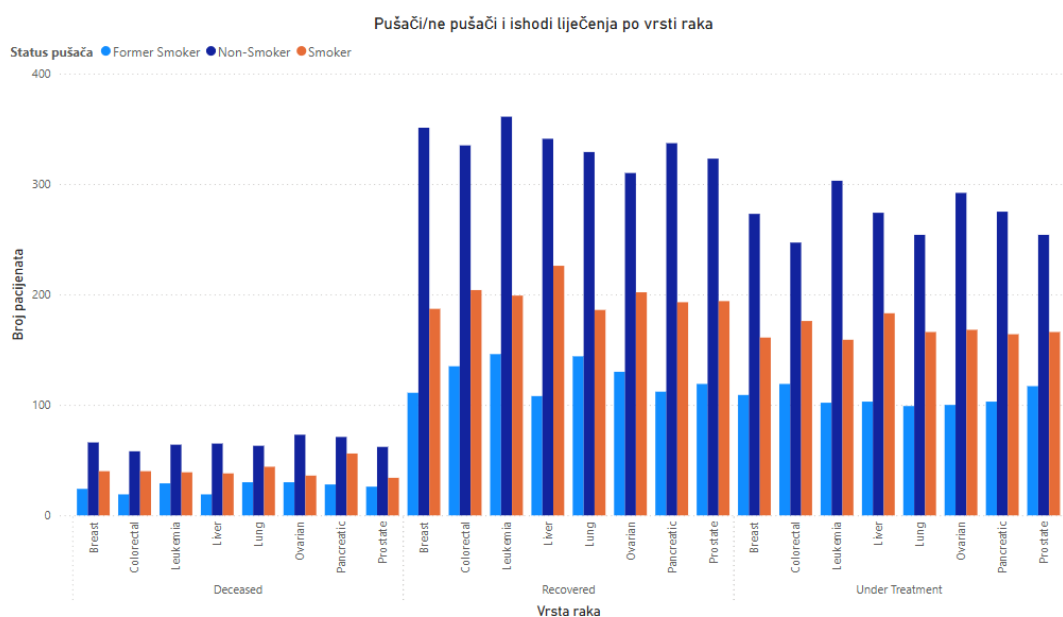
Ovaj izvještaj prikazuje raspodjelu uzoraka smrti među pacijentima. Najveći udio čine slučajevi kod kojih uzrok nije naveden, odnosno preživjeli su 90%. Pacijenti koji su preminuli i uzrok su bile komplikacije iznosi 5%, dok pacijenti koji su preminuli zbog karcinoma iznosi 4,78%.



Slika 29: Izvještaj o uzorku smrti, snimka zaslona(2025.)

4.1.6. Izvještaj o pušačima/ne pušačima i ishodu liječenja

Ovo izvješće prikazuje odnos između statusa pušenja (bivši pušači, nepušači i pušači) i ishoda liječenja po vrstama raka. Za sve vrste karcinoma i ishode liječenja (preminuli, izliječeni, na liječenju), najveći broj pacijenata čine nepušači, dok su pušači i bivši pušači zastupljeni u manjem broju. Ovakva raspodjela nam ukazuje da je među analiziranim pacijentima udio nepušača najviši bez obzira na vrstu raka ili ishod liječenja.



Slika 30: Izvješće o pušačima/ne pušačima i ishodu liječenja po vrsti raka, snimka zaslona(2025.)

5. Zaključak

Provedenom analizom uspješno su objedinjeni, strukturirani i vizualizirani podaci o onkološkim bolesnicima u Ujedinjenim Arapskim Emiratima, čime je omogućen pregled ključnih demografskih i kliničkih pokazatelja. Korištenjem skladišta podataka i alata za poslovnu inteligenciju, identificirani su najčešći tipovi karcinoma, ishodi liječenja po bolnicama i dobnim skupinama, koji su najčešći tipovi karcinoma te prostorna distribucija zdravstvenih ustanova. Dobiveni rezultati potvrđuju vrijednost kvalitete integracije i analize medicinskih podataka za podršku donošenja odluka i unapređenje zdravstvene skrbi. Izgrađeni sustav može poslužiti kao temelj za daljnja istraživanja i složenije analize u području onkologije.

Popis literature

Kaggle. (n.d.). *UAE Cancer Patient Dataset*. Preuzeto 11. 4. 2025. s <https://www.kaggle.com/datasets/ak0212/uae-cancer-patient-dataset/data>

Oracle (n.d.). *MySQL Workbench*. MySQL. Preuzeto 3. 6. 2025. s <https://www.mysql.com/products/workbench/>

Microsoft (n.d.). *Power BI*. Preuzeto 3.6. 2025. s <https://www.microsoft.com/en-us/power-platform/products/power-bi#Resources>

Microsoft Community (2020.). *How to Connect to an ODBC Data Source From Power BI*. Preuzeto 1. 6. 2025. s <https://community.fabric.microsoft.com/t5/Power-BI-Community-Blog/How-to-Connect-to-an-ODBC-Data-Source-From-Power-BI/ba-p/960701>

Oracle (n.d.) *MySQL :: Download MySQL Connector/ODBC*. Preuzeto 1. 6. 2025. s <https://dev.mysql.com/downloads/connector/odbc/>

Microsoft Learn (n.d.). *DAX function refrence*. Preuzeto 1. 6. 2025 s <https://learn.microsoft.com/en-us/dax/dax-function-reference>

Popis slika

Slika 1: Prikaz podataka u Excel-u	4
Slika 2: Kreiranje lokalnog servera	5
Slika 3: Kreiranje i testiranje konekcije za lokalni poslužitelj	5
Slika 4: Kreiranje sheme cancer_dw	5
Slika 5: Kreiranje privremene tablice za uvoz podataka	6
Slika 6: Uvoz podataka pomoću čarobnjaka	6
Slika 7: Uvoz podataka u tablicu	7
Slika 8: Prikaz uvezenih podataka.....	7
Slika 9: Kreiranje tablice patient	8
Slika 10: Unos podataka u tablicu patient iz staging_table	8
Slika 11: Kreiranje tablice date_dim	9
Slika 12: Kreiranje tablice cancer	9
Slika 13: Kreiranje tablice location.....	9
Slika 14: Kreiranje tablice treatment.....	9
Slika 15: Unos podataka u tablicu cancer iz staging_table	10
Slika 16: Unos podataka u tablicu dim_date iz staging_table	10
Slika 17: Unos podataka u tablicu location iz staging_table.....	10
Slika 18: Unos podataka u tablicu treatment iz staging_table	11
Slika 19: ERA model	11
Slika 20: Odabir dohvaćanja podataka	12
Slika 21: Odabir dohvaćanja podataka (vrsta povezivanja)	13
Slika 22: Odabir izvora podataka.....	13
Slika 23: Odabir tablica za učitavanje u Power BI	14
Slika 24: Početno stanje nakon učitavanja tablica	14
Slika 25: Izvještaj o broju slučajeva po dobnim skupinama i vrsti raka	16
Slika 26: Izvještaj o ishodima liječenja po bolnicama.....	17
Slika 27: Izvještaj o lokacijama bolnica.....	18
Slika 28: Izvještaj o ishodima tretmana pacijenata po vrsti karcinoma	19
Slika 29: Izvještaj o uzorku smrti	19
Slika 30: Izvješće o pušačima/ne pušačima i ishodu liječenja po vrsti raka	20