

Rezervacija Apartmana

- Dokumentacija –

Autor: Aleksa Milenović E7 2/2022

Asistent: Vladimir Ivković

Profesor: vandr. prof. dr. Slavica Kordić

Sadržaj

1. Motivacija i ciljevi	3
2. Opis zadataka i plan realizacije projekta	3
3. Arhitektura Data Warehouse sistema.....	4
4. Opis OLTP baze podataka i drugih izvora podataka.....	4
4.1 Izvori podataka.....	4
4.2 OLTP baze podataka	5
5. Opis OLAP baze podataka	9
5.1 Dimenzione tabele	10
5.2 Tabele činjenica	12
6. Opis ECTL procesa	14
7. Dobijeni rezultati i izveštaji.....	18
7.1 Izveštaji za prvu temu poslovanja	18
7.2 Izveštaji za drugu temu poslovanja	20
7.3 Izveštaji za treću temu poslovanja	21
8. Zaključak	23

1. Motivacija i ciljevi

Postoji veoma velik značaj podataka u ugostiteljskoj industriji. Kako bi se donosile informisane poslovne odluke i poboljšalo iskustvo korisnika, od suštinskog je značaja imati tačne i pouzdane podatke. Sistem za skladištenje podataka (*eng. Data Warehouse*) može pomoći u rešavanju izazova upravljanja i integrisanja podataka sa različitih izvora, ali i obezbediti ujedineni prikaz podataka koji se može koristiti za izveštavanje, analizu i donošenje odluka.

Cilj ovog projekta jeste pružiti oglašavaču apartmana uvid u njegovo poslovanje, tj. poslovanje izdavanja smeštaja, koje će biti pokriveno kroz tri glavne teme poslovanja. Na osnovu njih on će biti u mogućnosti da donese niz odluka u cilju unapređenja svog načina rada zarad ostvarivanja veće profitabilnosti. Ove teme poslovanja su:

1. Praćenje priliva novca i broja noćenja za svaki apartman u odnosu na način oglašavanja (Booking, Airbnb ili OLTP baza podataka). Ovde je uključeno i praćenje broja noćenja u zavisnosti od porekla gostiju.
2. Druga tema treba da podrži donošenje odluke da li je isplativo da za određeni broj noćenja postoji različitu cenovnu politiku. Kao glavni faktor na ovo utiče period godine ili nedelje, gde se posmatrati popunjenost tj. nepopunjenost smeštajnih kapaciteta. Pored ovoga gleda se i dužina samih rezervacija u danima.
3. Treća tema koju ovaj sistem treba da podrži jeste donošenje odluke na koji način i u kojoj meri sprovesti politiku otkazivanja rezervacija (*eng. cancellation policy*), tj. donošenje odluke koliko unapred klijent može da otkáže rezervaciju bez posledice (neke novčane kompenzacije) ali isto tako i u zavisnosti od dužine boravka, koliko je dana unapred rezervacija napravljena i same cene aranžmana.

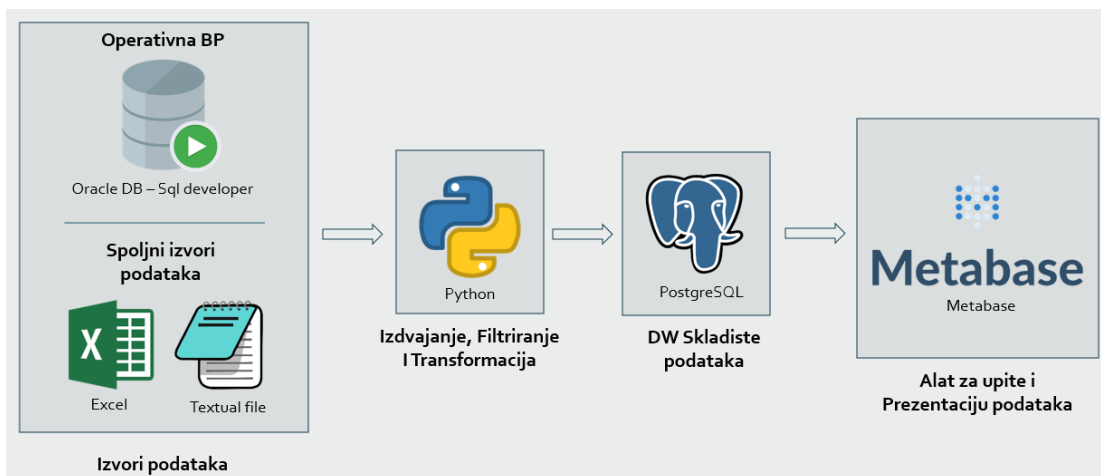
2. Opis zadataka i plan realizacije projekta

Zadatak ovog projekta je razviti data warehouse relacionu bazu podataka za sistem rezervacije apartmana. Na osnovu napravljenog data warehouse-a razvijaju se izveštaji o podacima, koji će naručiocu projekta pružiti uvid u pravilnosti i poslovanje njegovog biznisa. Plan realizacije projekta je podeljen u celina:

1. Sagledanje potreba korisnika i poslovanja
2. Prikupljanje i čišćenje podataka
3. Projektovanje OLTP šeme baze podataka na osnovu SQL skripti
4. Projektovanje OLAP šeme baze podataka na osnovu SQL skripti
5. Konstrukcija ECTL procesa
6. Generisanje izveštaja

3. Arhitektura Data Warehouse sistema

Projekat je realizovan korišćenjem programskih jezika *Python* i *SQL*. Prvu celinu arhitekture čine izvori podataka koji detaljnije opisani u narednom poglavlju. U ove izvore izvori podataka spadaju excel i tekstualni fajlovi, kao operativna baza podataka koja je pomoću SQL skripte formirana u *Oracle SQL Developer-u*. Izvori podataka se zatim učitavaju u *Jupyter Notebook*, gde se korišćenjem *Pandas* biblioteke nad podacima primenjuju različite transformacije, nakon čega se oni učitavaju u *PostgreSQL* bazu podataka gde je izlazna *Data Warehouse* šema smeštena. *PostgreSQL* se zatim povezuje sa *Metabase* programom koji predstavlja alat za vizualizaciju, gde se korišćenjem SQL upita formiraju izveštaji prikazani u poslednjem poglavlju.



Slika 3.1 Prikaz arhitekture sistema

4. Opis OLTP baze podataka i drugih izvora podataka

U ovom poglavlju su predstavljeni različiti izvori podataka kao i opis inicijalne OLTP baze podataka sa svojim tabelama, obeležjima i ograničenjima.

4.1 Izvori podataka

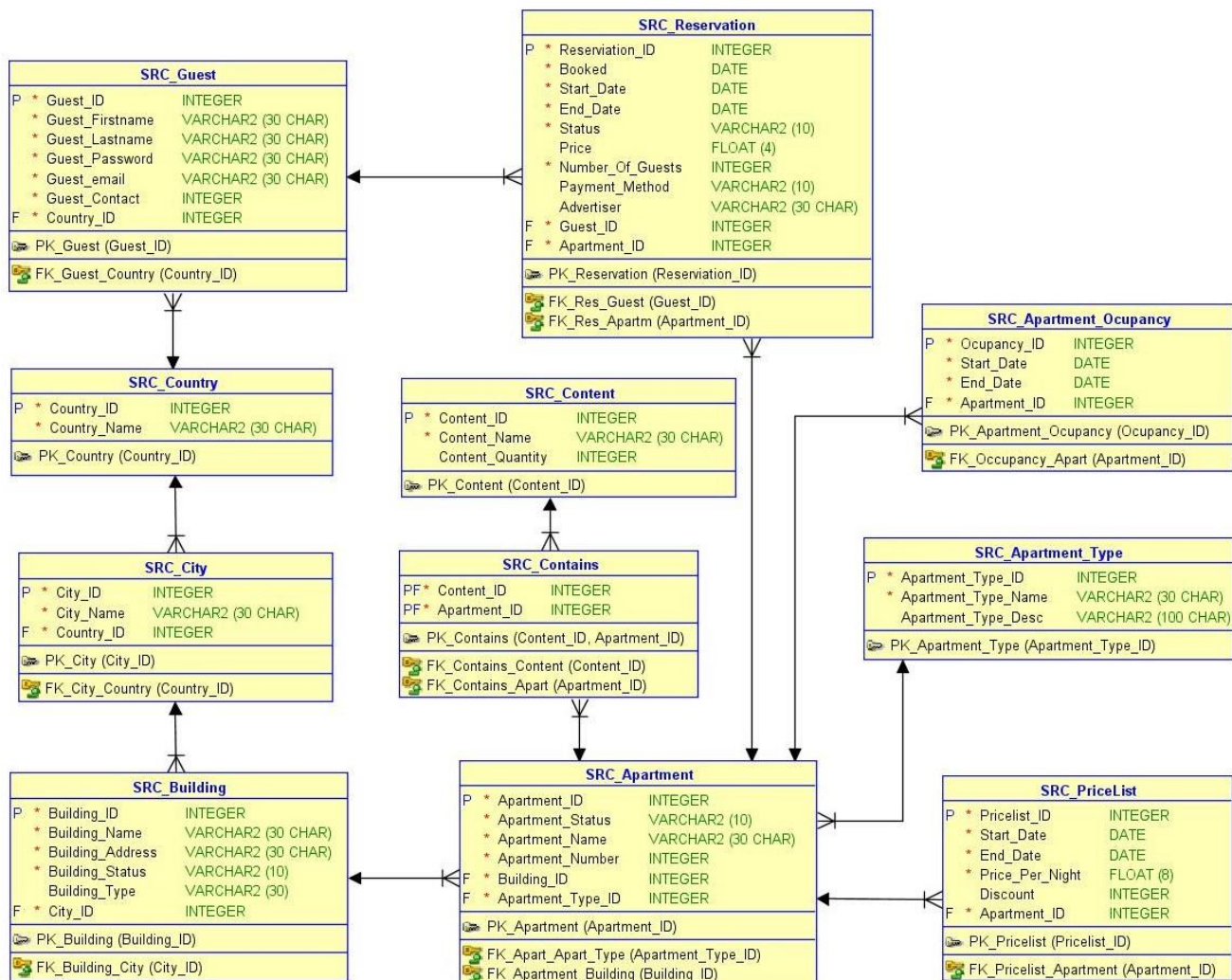
Kako bi se uspešno izvršile analize i kreiranje izveštaja, najpre je neophodno prikupiti sve potrebne podatke. Izvori podataka u bazi su sledeći:

1. Kao prvi izvor podataka korišćene su datoteke koje obuhvataju podatke o rezervacijama napravljenim na Booking internet stranici (www.booking.com). Podaci su dati u vidu 4 excel fajla (.xlsx), gde svaki fajl odgovara jednoj od godini između 2018. i 2022. godine. Usled nedostatka u podacima, za potrebe ovog projekta korišćeni su samo podaci iz 2021. i 2022. godine. Postoji 148 rezervacija koje su opisane sa 11 obeležja u koje spadaju ime i prezime gosta, datum početka i kraja rezervacije, kao i neki informativni podaci o samim smeštajima.
2. Kao drugi izvor podataka korišćene su datoteke koje obuhvataju podatke o rezervacijama napravljenim na Airbnb internet stranici (www.airbnb.com). Podaci su takođe dati u vidu 4 excel fajla (.xlsx), gde svaki fajl odgovara jednoj od godina između 2018. i 2022. godine. Postoji 108 rezervacija koje su opisane sa po 13 obeležja, i važno je napomenuti da je u cenu uračunata i boravišna taksa koja iznosi 1 eur po osobi dnevno, dok se na Booking-u taksa plaća odvojeno.

- Uz excel fajlove priložen je i jedan tekstualan fajl (.txt) koji od podataka sadrži naziv apartmana, broj kreveta ali i šifru pod kojom se apartman vodi u excel datotekama.
- Kao poslednji izvor podataka, za ovaj Data Warehouse sitem, iskorišćen je deo šeme baze podataka sistema za rezervisanje smeštaja, koji vodi evidenciju o izvršenim rezervacijama za apartmane. Ova šema kreirana je za potrebe ovog projekta, ali se može koristiti kao osnova za pravljenje nove internet aplikacije za oglašavanje i rezervaciju smeštaja. Model se sastoji od 11 tabela. Podaci su veštački simulirani, i njihova količina je zanemarljivo mala, naspram one koja bi bila u realnim uslovima.

4.2 OLTP baze podataka

Na slici 4.1, prikazana je logička sema OLTP (*On-Line Transaction Processing*) baze podataka, koja predstavlja poslednji izvor podataka predstavljen u prethodnom poglavlju. U daljem tekstu dat je opis svake od tabela iz OLTP šeme baze podataka, pri čemu je za svaku od tabela predstavljen skup njenih obeležja, dat opis za svako obeležje i opisani primarni i strani ključevi.



Slika 4.1 Logička OLTP sema baze podataka

Tabela Gost (SRC_Guest)

Tabela SRC_Guest sadrži podatke o gostu koji je registrovan na sistem.

Tabela 4.1.1 – Kolone u tabeli SRC_Guest

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Guest_ID	Long Integer	Ne	Jedinstveni identifikator gosta
2.	Guest_Firstname	Varchar2(30)	Ne	Ime gosta
3.	Guest_Lastname	Varchar2(30)	Ne	Prezime gosta
4.	Guest_Password	Varchar2(30)	Ne	Lozinka gostovog naloga
5.	Guest_email	Varchar2(30)	Ne	Mail gosta
6.	Guest_Contact	Integer	Da	Broj telefona gosta

Ograničenja:

- PK_Guest (Guest_ID)
- FK_Guest_Country (Country_ID): SRC_Guest [Country_ID] \subseteq SRC_Country [Country_ID]

Tabela Objekat (SRC_Building)

Tabela SRC_Building sadrži podatke o objektu unutar kog se nalaze apartmani koji se izdaju.

Tabela 4.1.2 – Kolone u tabeli SRC_Object

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Building_ID	Long Integer	Ne	Jedinstveni identifikator objekta
2.	Building_Name	Varchar2(30)	Ne	Naziv objekta
3.	Building_Address	Varchar2(100)	Ne	Adresa na kojoj se objekat nalazi
4.	Building_Status	STATUS_OBJ	Ne	Trenutan status objekta
5.	Building_Type	TIP_OBJ	Ne	Tip kojem objekat pripada

Ograničenja:

- PK_Building (Building_ID)
- FK_Building_City (City_ID): SRC_Building [City_ID] \subseteq SRC_City [City_ID]

Tabela Zauzetost Apartmana (SRC_Apartment_Ocupancy)

Tabela SRC_Apartment_Ocupancy sadrži podatke o datumima koji su rezervisani od strane gosta za neki apartman.

Tabela 4.1.3 – Kolone u tabeli SRC_Apartment_Ocupancy

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Ocupancy_ID	Long Integer	Ne	Jedinstveni identifikator zauzetosti apartmana
2.	Start_Date	Date & Time	Ne	Pocetni datum rezervacije
3.	End_Date	Date & Time	Ne	Krajnji datum te rezervacije

Ograničenja:

- PK_Apartment_Ocupancy (Ocupancy_ID)
- FK_Ocupancy_Apart (Apartment_ID): SRC_Apartment_Ocupancy[Apartment_ID] \subseteq SRC_Apartment[Apartment_ID]

Tabela Apartman (SRC_Apartment)

Tabela SRC_Apartment sadrži podatke o apartmanu koji vlasnik izdaje.

Tabela 4.1.4 – Kolone u tabeli SRC_Apartment

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Apartment_ID	Long Integer	Ne	Jedinstveni identifikator apartmana
2.	Apartment_Name	Varchar2(30)	Ne	Naziv apartmana
3.	Apartment_Status	STATUS_APRT	Ne	Status apartmana
4.	Apartment_Number	Integer	Ne	Broj apartmana

Ograničenja:

- PK_Apartment (Apartment_ID)
- FK_Apart_Apart_Type (Aprt_Type_ID): SRC_Apartment [Aprt_Type_ID] \subseteq SRC_Apartment_Type [Aprt_Type_ID]
- FK_Apart_Building (Building_ID): SRC_Apartment [Building_ID] \subseteq SRC_Building [Building_ID]

Tabela Tip Apartmana (SRC_Apartment_Type)

Tabela SRC_Apartment_Type sadrži podatke o tipu tj. kategoriji kojoj apartman može da pripada.

Tabela 4.1.5 – Kolone u tabeli SRC_Apartment_Type

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Aprt_Type_ID	Long Integer	Ne	Jedinstveni identifikator tipa apartmana
2.	Aprt_Type_Name	Varchar2(30)	Ne	Naziv tipa apartmana
3.	Aprt_Type_Desc	Varchar2(100)	Da	Opis tog tipa apartmana

Ograničenja:

- PK_Apartment_Type (Arpt_Type_ID)

Tabela Cenovnik (SRC_Pricelist)

Tabela SRC_Pricelist sadrži podatke o cenama noćenja u apartmanu koje važe za neki određen vremenski period.

Tabela 4.1.6 – Kolone u tabeli SRC_Pricelist

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Pricelist_ID	Long Integer	Ne	Jedinstveni identifikator cenovnika
2.	Start_Date	Date & Time	Ne	Pocetni datum vazenja cenovnika
3.	End_Date	Date & Time	Ne	Krajnji datum vazenja cenovnika
4.	Price_Per_Night	Float	Ne	Cena po nocenju u apartmanu
5.	Discount	Integer	Da	Popust na cenu noci u apartmanu

Ograničenja:

- PK_Pricelist (Pricelist_ID)
- FK_Pricelist_Apartment (Apartment_ID): SRC_Pricelist [Apartment_ID] \subseteq SRC_Apartment [Apartment_ID]

Tabela Sadržaj (SRC_Content)

Tabela SRC_Content sadrži podatke o svom sadržaju koji apartman može da poseduje (npr. br. kreveta 3)

Tabela 4.1.7 – Kolone u tabeli SRC_Content

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Content_ID	Long Integer	Ne	Jedinstveni identifikator sadržaja apartmana
2.	Content_Name	Varchar2(30)	Ne	Naziv sadržaja koji apartman poseduje
3.	Content_Quantity	Integer	Da	Kolicina sadržaja koji apartman poseduje

Ograničenja:

- PK_Content (Content_ID)

Tabela Rezervacija (SRC_Reservations)

Tabela SRC_Reservations sadrži podatke o rezervaciji apartmana koju gost pravi.

Tabela 4.1.8 – Kolone u tabeli SRC_Reservations

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Reservations_ID	Long Integer	Ne	Jedinstveni identifikator rezeravacije
2.	Booked	Date & Time	Ne	Datum kreiranja rezervacije
3.	Start_Date	Date & Time	Ne	Pocetni datum rezervacije, dolaska
4.	End_Date	Date & Time	Ne	Krajnji datum rezervacije, odlaska
5.	Status	STATUS_REZ	Ne	Status rezervacije
6.	Price	Float	Ne	Ukupna cena rezervacije koju gost treba da plati
7.	Number_Of_Guests	Integer	Ne	Ukupan broj ljudi koji ce odседati u apartmanu
8.	Payment_Method	NACIN_PL	Ne	Nacin placanja rezervacije

Ograničenja:

- PK_Reservations (Reservations_ID)
- FK_Res_Guest (Guest_ID): SRC_Reservations [Guest_ID] \subseteq SRC_Guest [Guest_ID]
- FK_Res_Apart (Apartment_ID): SRC_Reservations [Apartment_ID] \subseteq SRC_Apartment [Apartment_ID]

Tabela Sadrzi (Src_Contains)

Tabela Recenzija sadrži podatke recenzijama koju gost može da ostavi za apartman u kom je prethodno odseo.

Tabela 4.1.9 – Kolone u tabeli Recenzija

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Content_ID	Long Integer	Ne	Jedinstveni identifikator sadržaja
2.	Apartment_ID	Long Integer	Ne	Jedinstveni identifikator apartmana

Ograničenja:

- PK_Contains (Content_ID + Apartment_ID)
- FK_Contains_Content (Content_ID): SRC_Contains [Content_ID] \subseteq SRC_Content [Content_ID]
- FK_Contains_Aprt (Apartment_ID): SRC_Contains [Apartment_ID] \subseteq SRC_Apartment [Apartment_ID]

Tabela Grad (SRC_City)

Tabela SRC_City sadrži podatke o gradovima država u kojima se nalaze apartmani.

Tabela 4.1.10 – Kolone u tabeli SRC_City

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	City_ID	Long Integer	Ne	Jedinstveni identifikator grada
2.	City_Name	Varchar2(30)	Ne	Naziv grada

Ograničenja:

- PK_City (City_ID)
- FK_City_Country (Country_ID): SRC_City [Country_ID] \subseteq SRC_Country [Country_ID]

Tabela Drzava (SRC_Country)

Tabela SRC_Country sadrži podatke o državama u kojima se nalaze apartmani, kao i države porekla gostiju.

Tabela 4.1.11 – Kolone u tabeli SRC_Country

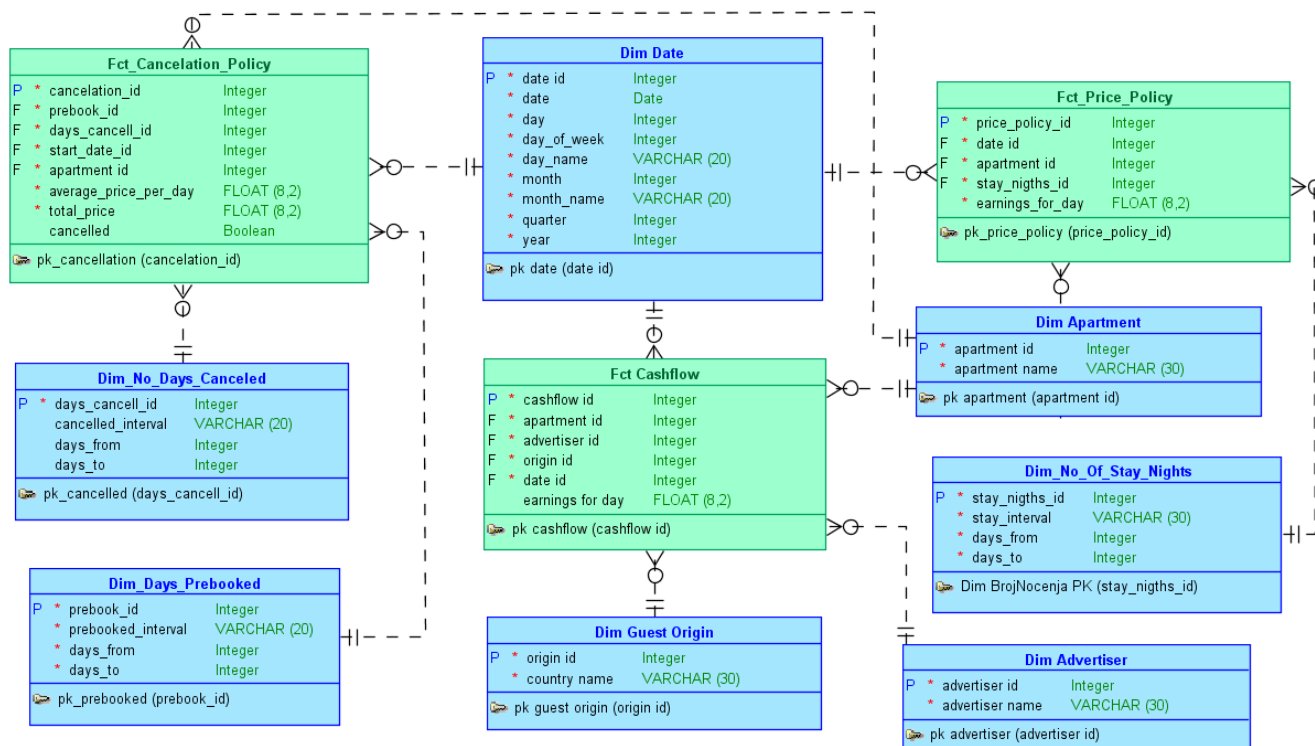
Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Country_ID	Long Integer	Ne	Jedinstveni identifikator drzave
2.	Country_Name	Varchar2(30)	Ne	Naziv drzave

Ograničenja:

- PK_Country (Country_ID)

5. Opis OLAP baze podataka

Na slici 4.1 je prikazana konceptualna šema OLAP baze podataka, tj. šema baze Data Warehouse sistema, koja je dobijena nizom transformacija i agregacija podataka iz OLTP baze podataka kao i excel fajlova booking i airbnb sajtova.



Slika 5.1 Konceptualna OLTP sema baze podataka

U daljem tekstu dat je opis svake od tabela iz OLTP šeme baze podataka. Šema je modelovana u obliku sazvežđa sa 7 dimenzionih tabela i 3 tabele činjenica, pri čemu je za svaku od tabela predstavljen skup njenih obeležja, dat opis za svaku od njih i opisani su primarni i strani ključevi. Prvo su opisane dimenzione tabele (na slici plave boje), a nakon toga tabele činjenica (na slici zelene boje).

5.1 Dimenzione tabele

Dimenzione tabele sadrže podatke koji opisuju, tj. daju kontekst podacima u tabelama činjenica. U ovoj šemi to su tabele za: Vreme, Apartman, Broja Noćenja, Oglašavača, Broj Dana Unapred Otkazano, Broj Dana Unapred Rezervisano i Porekla Gosta.

Tabela Apartmana (Dim_Apartment)

Tabela Dim_Apartment predstavlja dimenziju koja čuva podatke o nazivima apartmana. Podaci se dobavljaju iz tabele SRC_Apartment izvorne baze podataka, i polja naziv iz ApartmentTypes.txt fajla. Nivo je nivo pojedinačnog naziva apartmana. Tabela je jedno-hijerarhijske dimenzije.

Tabela 5.1.1 – Kolone u tabeli Dim_Apartment

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Apartment_ID	Long Integer	Ne	Jedinstveni identifikator apartmana
2.	Apartment_Name	Varchar2(30)	Ne	Naziv apartmana

Ograničenja:

- PK_Apartment (Apartment_ID)

Tabela Oglašavača (Dim_Advertiser)

Tabela Dim_Advertiser predstavlja dimenziju koja čuva podatke o nazivima oglašavača. Nivo je nivo jednog pojedinačnog oglašavača. Tabela je jedno-hijerarhijske dimenzije.

Tabela 5.1.2 – Kolone u tabeli Dim_Advertiser

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Advertiser_ID	Long Integer	Ne	Jedinstveni identifikator oglasavaca
2.	Advertiser_Name	Varchar2(30)	Ne	Naziv oglasavaca

Ograničenja:

- PK_Advertiser (Advertiser_ID)

Tabela Porekla Gosta (Dim_Guest_Origin)

Tabela Dim_Guest_Origin predstavlja dimenziju koja čuva podatke o poreklu gostiju koji su napravili rezervaciju za neki od apartmana. Podaci se dobavljaju iz tabela SRC_Guest i SRC_Country izvorne baze podataka, dok se iz *booking* i *airbnb* excel fajlova dobijaju korišćenjem python biblioteke *NameDataset* za određivanje porekla na osnovu imena i prezimena gosta. Nivo je nivo pojedinačnog porekla gosta. Tabela je jedno-hijerarhijske dimenzije.

Tabela 5.1.3 – Kolone u tabeli Dim_Guest_Origin

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Origin_ID	Long Integer	Ne	Jedinstveni identifikator tipa gosta
2.	Country_Name	Varchar2(30)	Ne	Drzava iz koje su gosti jedne rezervacije

Ograničenja:

- PK_Guest_Origin (Origin_ID)

Tabela Broja Unapred Rezervisanih dana (Dim_Days_Prebooked)

Tabela Dim_Days_Prebooked predstavlja intervalnu dimenziju koja čuva podatke o broju dana koliko je unapred gost napravio rezervaciju, tj razliku između početnog datuma aranžmana i datuma pravljenja rezervacije. Podaci se određuju spram karakterističnih opsega broja dana. Nivo je nivo broja dana. Tabela je jedno-hijerarhijske dimenzije.

Tabela 5.1.4 – Kolone u tabeli Dim_Days_Prebooked

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Prebook_ID	Long Integer	Ne	Jedinstveni identifikator broja dana
2.	Days_From	Integer	Ne	Donja granica intervala ukupnog broja dana
3.	Days_To	Integer	Ne	Gornja granica intervala ukupnog broja dana
4.	Prebook_Interval	Varchar2(30)	Ne	Tekstualni opseg broja dana

Ograničenja:

- PK_Prebooked (Prebook_ID)

Tabela Broja Dana Unapred Otkazivanja (Dim_Days_Cancelled)

Tabela Dim_Days_Cancelled predstavlja intervalnu dimenziju koja čuva podatke o broju dana koliko je unapred gost otkazao rezervaciju, tj razliku između početnog datuma aranžmana i datuma otkazivanja rezervacije. Podaci se određuju spram karakterističnih opsega broja dana. Nivo je nivo broja dana. Tabela je jedno-hijerarhijske dimenzije.

Tabela 5.1.5 – Kolone u tabeli Dim_Days_Cancelled

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Days_Cancel_ID	Long Integer	Ne	Jedinstveni identifikator broja dana
2.	Days_From	Integer	Ne	Donja granica intervala ukupnog broja dana
3.	Days_To	Integer	Ne	Gornja granica intervala ukupnog broja dana
4.	Cancelled_Interval	Varchar2(30)	Ne	Tekstualni opseg broja dana

Ograničenja:

- PK_Canceled (Days_Cancel_ID)

Tabela Broja Noćenja (Dim_No_Of_Stay_Nights)

Tabela Dim_No_Of_Stay_Nights predstavlja intervalnu dimenziju koja čuva podatke ukupnom broju noćenja u okviru jedne rezervacije. Omogućava praćenje broja noćenja. Podaci se dobivljaju iz tabele REZERVACIJA izvorne baze podataka. Nivo je nivo broja noćenja. Tabela je jedno-hijerarhijske dimenzije.

Tabela 5.1.6 – Kolone u tabeli Dim_No_Of_Stay_Nights

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Stay_Nights_ID	Long Integer	Ne	Jedinstveni identifikator broja noćenja
2.	Days_From	Integer	Ne	Donja granica intervala ukupnog broja noćenja
3.	Days_To	Integer	Ne	Gornja granica intervala ukupnog broja noćenja
4.	Stay_Interval	Varchar2(30)	Ne	Tekstualni opseg broja noćenja

Ograničenja:

- PK_Stay_Nights (Stay_Nights_ID)

Tabela Vremena (Dim_Date)

Tabela Dim_Date predstavlja obaveznu dimenziju kako bi bilo omogućeno vremensko praćenje. Nivo je nivo pojedinačnog dana, kako bi se omogućilo preciznije ukupnjavanje upita prema preciznijim vremenskim opsezima. Tabela je jedno-hijerarhijske dimenzije.

Tabela 5.1.7 – Kolone u tabeli Dim_Date

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Date_ID	Long Integer	Ne	Jedinstveni identifikator
2.	Date	Date	Ne	Datum
3.	Day	Integer	Ne	Dan za predstavljeni datum
4.	Day_Of_Week	Integer	Ne	Redni broj dana u nedelji za predstavljeni datum
5.	Day_Name	Varchar2(30)	Ne	Ime dana u nedelji za predstavljeni datum
6.	Month	Integer	Ne	Redni broj mesec u godini za predstavljeni datum
7.	Month_Name	Varchar2(30)	Ne	Ime meseca za predstavljeni datum
8.	Quarter	Integer	Ne	Redni broj kvartala u godini za predstavljeni datum
9.	Year	Integer	Ne	Godina za predstavljeni datum

Ograničenja:

- PK_Date (Date_ID)

5.2 Tabele činjenica

Tabele činjenica sadrže podatke koji omogućavaju davanje odgovora na određene upite. U ovoj šemi to su tabele: Fct_Cashflow, Fct_Price_Policy i Fct_Cancelation_Policy.

Tabela Priliva Novca (Fct_Cashflow)

Tabela Fct_Cashflow beleži priliv novca na nivou dana, pojedinačnog apartmana za svakog od oglašivača kao i za svako poreklo gostiju. Ovde se svaka realizovana rezervacija razdvaja na pojedinačne dane (npr. rezervacija od 3 dana će činiti tri reda pojedinačnih dana) kako bi se omogućilo preciznije postavljanje vremenskog opsega prilikom pretrage jer se sa manjeg može uvek agregirati na viši nivo. Povezana je sa tabelama dimenzija Dim_Date, Dim_Apartment, Dim_Guest-Origin, Dim_Advertiser. Njeno postojanje ukazuje

na događaj. Earnings_For_Day je aditivna mera jer je sabirljiva po svim dimenzijama. Svi atributi mera su osnovni jer se preuzimaju podaci iz OLTP baze podataka, kao i priloženih datoteka.

Tabela 5.2.1 – Kolone u tabeli Fct_Poslovanje

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Cashflow_ID	Long Integer	Ne	Jedinstveni identifikator priliva novca
2.	Apartment_ID	Long Integer	Ne	Jedinstveni identifikator apartmana
3.	Advertiser_ID	Long Integer	Ne	Jedinstveni identifikator oglašivača
4.	Origin_ID	Long Integer	Ne	Jedinstveni identifikator porekla gosta
5.	Date_ID	Long Integer	Ne	Jedinstveni identifikator datuma
6.	Earnings_For_Day	Float	Da	Naplaćena cena jednog noćenja u apartmanu u okviru jedne rezervacije za određeni dan, tj. datum

Ograničenja:

- PK_Cashflow (Cashflow_ID)
- FK_Cashflow_Apartment (Apartment_ID): Fct_Cashflow [Apartment_ID] \subseteq Dim_Apartment [Apartment_ID]
- FK_Cashflow_Advertiser (Advertiser_ID): Fct_Cashflow [Advertiser_ID] \subseteq Dim_Advertiser [Advertiser_ID]
- FK_Cashflow-Origin (Origin_ID): Fct_Cashflow [Origin_ID] \subseteq Dim-Origin [Origin_ID]
- FK_Cashflow_Date (Date_ID): Fct_Cashflow [Date_ID] \subseteq Dim_Date [Date_ID]

Tabela Cenovne Politike (Fct_Price_Policy)

Tabela Fct_Price_Policy čuva podatke o naplaćenim rezervacijama svakog apartmana. Ovde se takođe svaka rezervacija razdvaja na pojedinačne dane, kako bi se agregacijom obezbedio uvid u broj rezervisanih dana tj. nerezervisanih dana u toku nekog vremenskog perioda. Povezana je sa tabelama Dim_Date, Dim_Apartment, Dim_No_Of_Stay_Nights. Earnings_For_Day je aditivna mera jer je sabirljiva po svim dimenzijama. Svi atributi mera su osnovni jer se preuzimaju podaci iz OLTP baze podataka, kao i priloženih datoteka.

Tabela 5.2.2 – Kolone u tabeli Fct_Price_Policy

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Price_Policy_ID	Long Integer	Ne	Jedinstveni identifikator cenovne politike
2.	Apartment_ID	Long Integer	Ne	Jedinstveni identifikator apartmana
3.	Date_ID	Long Integer	Ne	Jedinstveni identifikator datuma
4.	Stay_Nights_ID	Long Integer	Ne	Jedinstveni identifikator broja noćenja
5.	Earnings_For_Day	Float	Da	Naplaćena cena jednog noćenja u apartmanu u okviru jedne rezervacije

Ograničenja:

- PK_Price_Policy (Price_Policy_ID)
- FK_Price_Policy_Apartment (Apartment_ID): Fct_Price_Policy [Apartment_ID] \subseteq Dim_Apartment [Apartment_ID]
- FK_Price_Policy_Date (Date_ID): Fct_Price_Policy [Date_ID] \subseteq Dim_Date [Date_ID]
- FK_Price_Policy_Stay_Nights (Stay_Nights_ID): Fct_Price_Policy [Stay_Nights_ID] \subseteq Dim_Stay_Nights [Stay_Nights_ID]

Tabela Politike Otkazivanja (Fct_Cancellation_Policy)

Tabela Fct_Cancellation_Policy čuva podatke o otkazanim rezervacijama, ali i neotkazanim rezervacijama. Obezbeđuje vlasniku apartmana da odredi politiku koliko dana unapred je moguće otkazati rezervaciju a da se pritom ne treba i dalje platiti apartman/dobiti povraćaj novca. Nivo je nivo pojedinačne rezervacije, gde neaditivna mera Cancelled ukazuje na to da li je rezervacija otkazana ili ne. Povezana je sa

tabelama dimenzija Dim_Date, Dim_Apartment, Dim_Days_Prebooked, Dim_No_Days_Cancelled. Total_Price i Avg_Price_For_Day su aditivne mere jer su sabirljive po svim dimenzijama. Svi atributi mera su osnovni jer se preuzimaju podaci iz OLTP baze podataka, kao i priloženih datoteka.

Tabela 5.2.3 – Kolone u tabeli Fct_Cancellation_Policy

Redni Br.	Naziv kolone	Tip podatka	Nullable	Opis kolone
1.	Cancellation_ID	Long Integer	Ne	Jedinstveni identifikator otkazne politike
2.	Prebook_ID	Long Integer	Ne	Jedinstveni identifikator dana unapred rezervisano
3.	Days_Cancell_ID	Long Integer	Ne	Jedinstveni identifikator dana otkazivanja rezervacije
4.	Start_Date_ID	Long Integer	Ne	Jedinstveni identifikator pocetnog datuma rezervacije
5.	Apartment_ID	Long Integer	Ne	Jedinstveni identifikator apartmana
6.	Cancelled	Boolean	Da	Ukazuje na to da li je rezervacija otkazana ili ne
7.	Total_Price	Float	Da	Ukupna cena cele rezervacije
8.	Avg_Price_For_Day	Float	Da	Prosečna cena jednog noćenja rezervacije

Ograničenja:

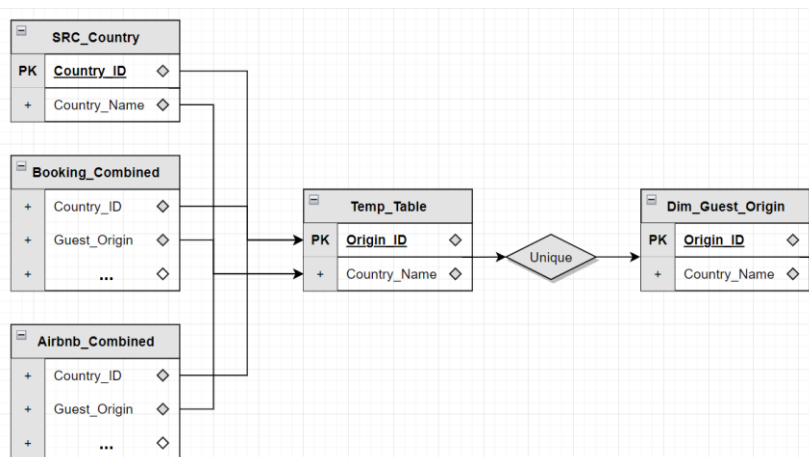
- PK_Cancellation_Policy (Cancellation_ID)
- FK_Cancel_Policy_Prebook (Prebook_ID): Fct_Cancellation_Policy [Prebook_ID] \subseteq Dim_Days_Prebooked [Prebook_ID]
- FK_Canc_Pol_Canc (Days_Cancel_ID): Fct_Cancellation_Pol [Days_Cancel_ID] \subseteq Dim_Days_Canceled [Days_Cancel_ID]
- FK_Cancel_Policy_Date (Start_Date_ID): Fct_Cancellation_Policy [Start_Date_ID] \subseteq Dim_Date [Start_Date_ID]
- FK_Cancel_Policy_Apart (Apartment_ID): Fct_Cancellation_Policy [Apartment_ID] \subseteq Dim_Apartment [Apartment_ID]

6. Opis ECTL procesa

U ovom poglavlju će biti opisan ECTL proces gde su prikazana mapiranja pomoću kojih su se punili podaci u OLAP bazu podataka. Prvo je opisano čišćenje podataka, nakon čega su opisana mapiranja dimenzionih tabela, pa tabela činjenica.

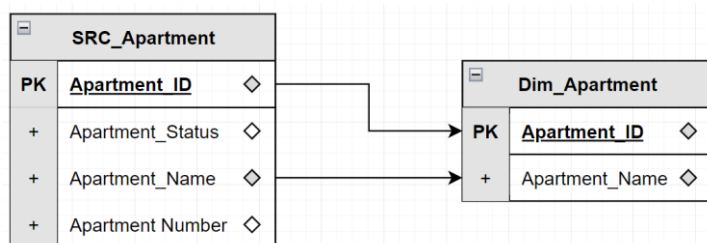
Pre samog ETL procesa obavljeno je čišćenje, formatiranje, ali i dodavanje neophodnih (dodatno preračunatih) podataka kod *Airbnb* i *Booking* excel fajlova. Kod oba fajla, korišćenjem *NameDataset* Python biblioteke određene su zemlje porekla klijenata na osnovu njihovih imena i prezimena, i dodate oznake zemalja. Pored ovoga izbačene su nepotrebne kolone.

Na slici 6.1 prikazano je mapiranje **Dim_Guest_Origin** tabele dimenzija koja prikazuje države gostiju.



Slika 6.1 – Mapiranje dimenzije Porekla Gostiju

Na slici 6.2 prikazano je mapiranje **Dim_Apartment** tabele dimenzija koja prikazuje nazive apartmana sa njihovim šiframa.



Slika 6.2 – Mapiranje tabele dimenzije Apartmana

Na slici 6.3 prikazana je kreirana **Dim_Advertiser** tabele dimenzija koja prikazuje nazive različitih oglašavača. Tabela je napravljeni ručno u programskom jeziku *Python* nakon čega će se u tabeli činjenica, u zavisnosti od porekla podataka, dodavati identifikacioni brojevi oglašavača.

advertiser_id	Advertiser
1	Airbnb
2	Booking
3	OLTP

Slika 6.3 – Kreiranje tabele dimenzije Oglašavača

Na slici 6.4 prikazana je kreirana **Dim_Length_Of_Stay** tabele dimenzija koja prikazuje različite vremenske intervale dužine rezervacija, tj. broja noćenja. Vremenski opsezi su napravljeni ručno u programskom jeziku *Python* nakon čega će se tabelama činjenica, preračunavanjem, dodavati identifikacioni brojevi ovih opsega.

length_of_stay_id	length_of_interval	date_from	date_to
0	1	1-2 days	1 2
1	2	3-5 days	3 5
2	3	6-10 days	6 10
3	4	10+ days	11 9999

Slika 6.4 – Kreiranje tabele dimenzije Dužine rezervacija

Na slici 6.5 prikazana je kreirana **Dim_Days_Prebooked** tabele dimenzija koja prikazuje različite vremenske intervale koji opisuju koliko dana unapred su rezervacije napravljene. Vremenski opsezi su napravljeni ručno u programskom jeziku *Python* nakon čega će se tabelama činjenica, preračunavanjem, dodavati identifikacioni brojevi ovih opsega.

prebook_id	prebook_interval	date_from	date_to
0	1	0-2 days	0 2
1	2	3-7 days	3 7
2	3	8-14 days	8 14
3	4	15-30 days	15 30
4	5	31-60 days	31 60

Slika 6.5 – Kreiranje tabele dimenzije Broja Dana unapred Rezervisano

Na slici 6.6 prikazana je kreirana **Dim_No_Days_Cancelled** tabele dimenzija koja prikazuje različite vremenske intervale koji opisuju koliko dana u odnosu na početni datum rezervacije je rezervacija otkazana. Vremenski opsezi su napravljeni ručno u programskom jeziku *Python* nakon čega će se tabelama činjenica, preračunavanjem, dodavati identifikacioni brojevi ovih opsega.

cancel_id	cancel_interval	date_from	date_to
0	1	0-2 days	0 2
1	2	3-7 days	3 7
2	3	8-14 days	8 14
3	4	15-30 days	15 30
4	5	31-60 days	31 60

Slika 6.6 – Kreiranje tabele dimenzije Broja Dana unapred Otkazano

Na slici 6.7 prikazano je kreiranje **Dim_Time** obavezne tabele dimenzija koja će sadržati sve datume u izabranom vremenskom periodu, zajedno sa identifikacionim brojem datuma, mesecom, godinom, nazivom dana u nedelji, nazivom meseca, itd.

```

dates = pd.date_range(start='2018-01-01', end='2023-12-31')

dim_Dates = pd.DataFrame({'date': dates, 'day': dates.day, 'month': dates.month, 'quarter': dates.quarter, 'year': dates.year})

dim_Dates['week_of_year'] = dim_Dates['date'].dt.isocalendar().week
dim_Dates['month_name'] = dim_Dates['date'].dt.strftime('%B')

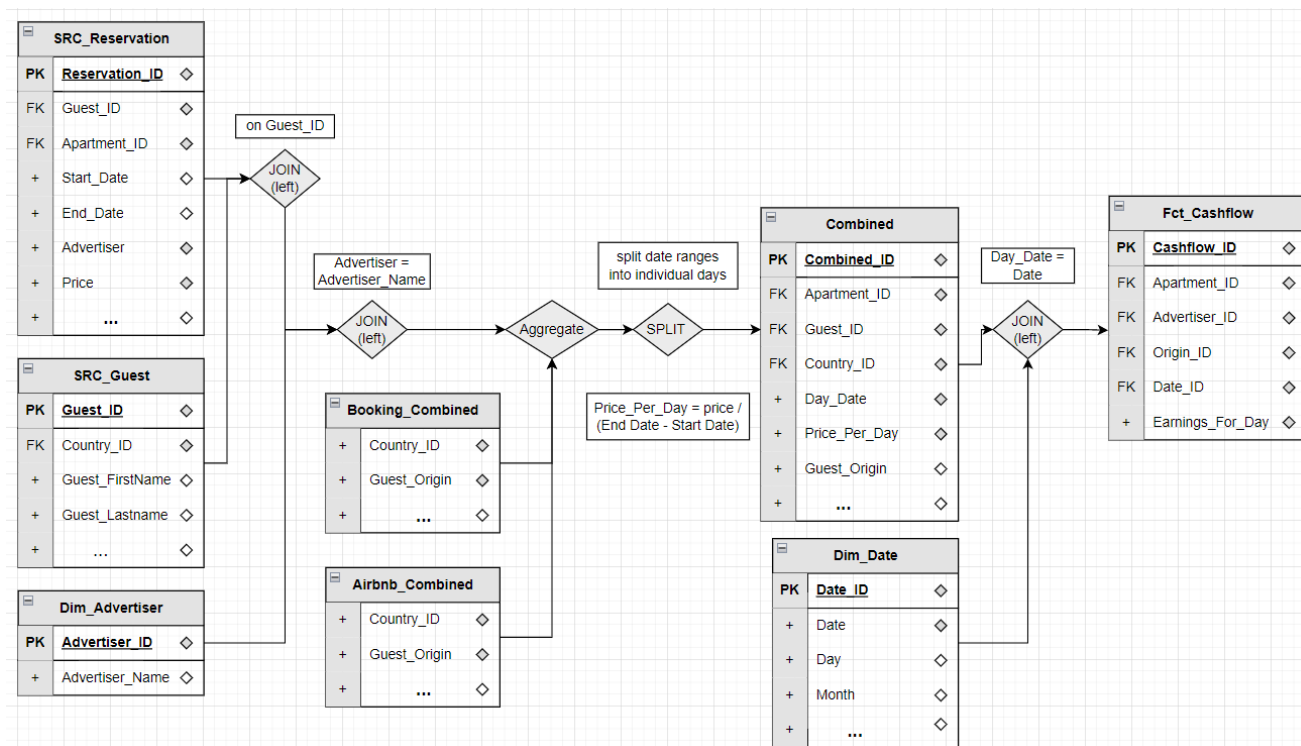
dim_Dates['day_of_week'] = dim_Dates['date'].dt.weekday
dim_Dates['day_name'] = dim_Dates['date'].dt.strftime('%A')

# set indexes
dim_Dates = dim_Dates.reset_index(drop = True)
dim_Dates['date_id'] = dim_Dates.index + 1

```

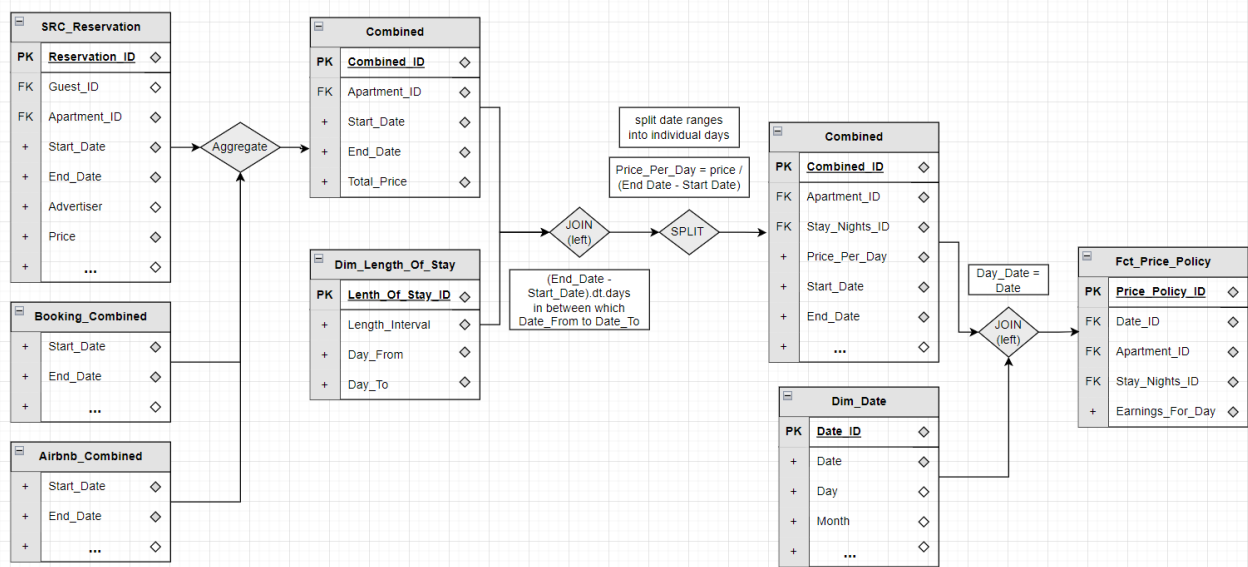
Slika 6.6 – Kreiranje tabele dimenzije Vremena

Na slici 6.7 prikazano je mapiranje **Fct_Cashflow** tabele činjenica koja se odnosi na prvu temu poslovanja. Rezervacije su nakon agregacije podeljene na nivoe pojedinačnih dana od početnog do krajnjeg datuma rezervacije pri čemu je ukupna cena podeljena sa brojem dana kako bi se dobila prosečna cena jednog noćenja.



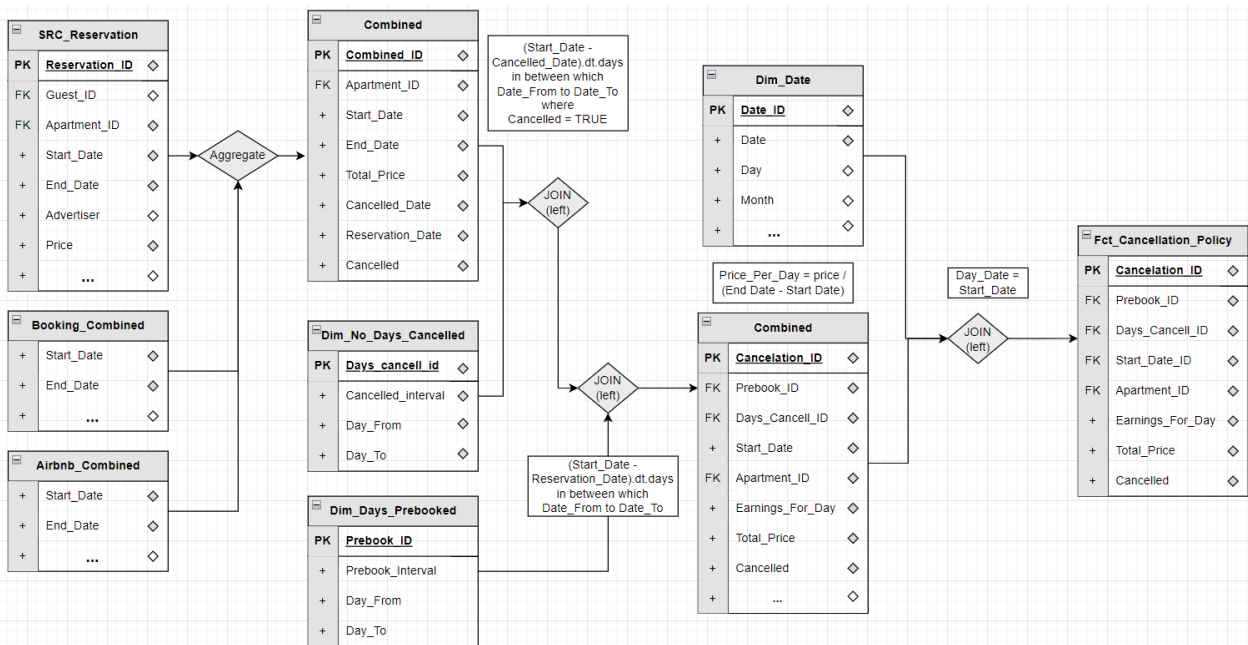
Slika 6.7 – Mapiranje tabele činjenica Cashflow

Na slici 6.8 prikazano je mapiranje **Fct_Price_Policy** tabele činjenica koja se odnosi na prvu drugu poslovanja. Rezervacije su nakon spajanja sa dimenzijom Length_Of_Stay podjeljene na nivoe pojedinačnih dana od početnog do krajnjeg datuma rezervacije pri čemu je ukupna cena podjeljena sa brojem dana kako bi se dobila prosečna cena jednog noćenja. Spajanje se dimenzijom Dužine Boravka je izvršeno na osnovu poređenja broja noćenja dobijenog iz razlike krajnjeg i prvog datuma rezervacije sa Day_From i Day_To atributima tabele dimenzija.



Slika 6.8 – Mapiranje tabele činjenica Price Policy

Na slici 6.9 prikazano je mapiranje **Fct_Cancellation_Policy** tabele činjenica koja se odnosi na prvu treću poslovanja. Nakon spajanja sa tabelama dimenzija ukupna cena je podjeljena sa brojem dana kako bi se dobila prosečna cena jednog noćenja. Spajanje se dimenzijom Unapred Otkazano je izvršeno na osnovu poređenja razlike početnog datuma i datuma otkazivanja sa Day_From i Day_To atributima tabele dimenzija, a isto je urađeno i za dimenzijom broja Dana Unapred Rezervisanog.



Slika 6.9 – Mapiranje tabele činjenica Cancellation Policy

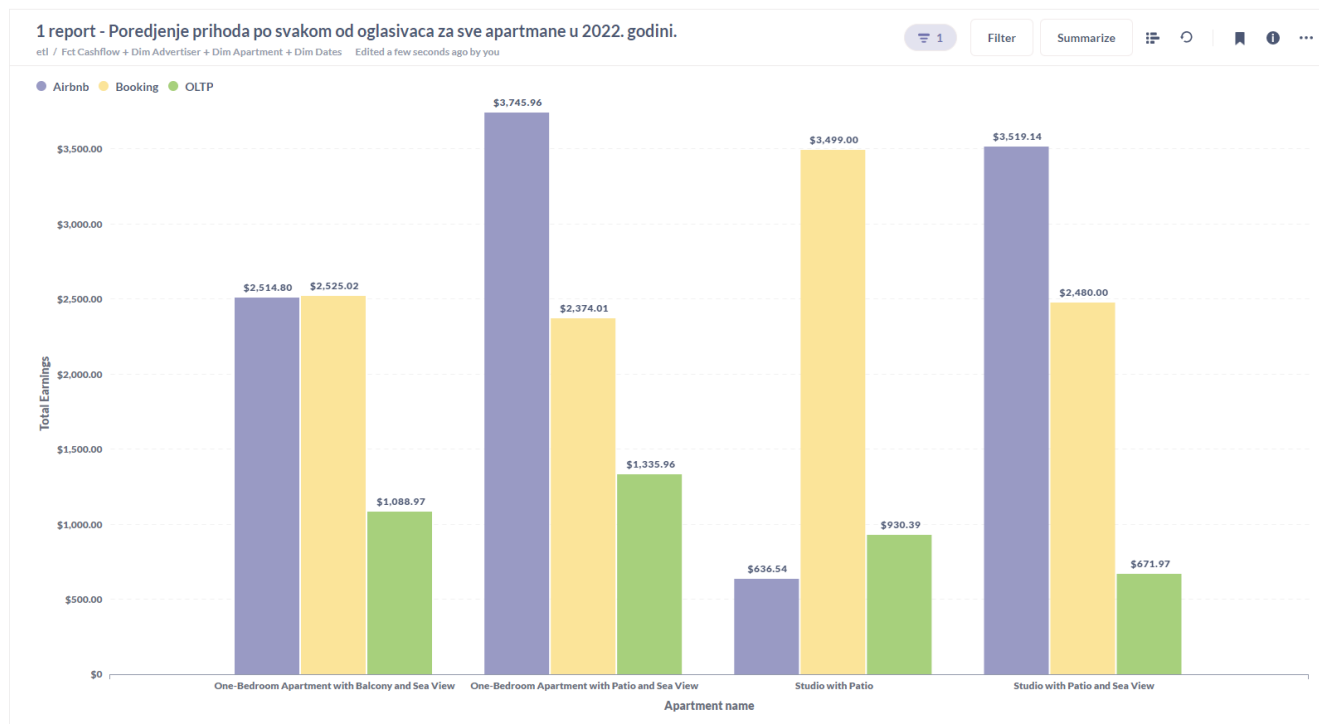
7. Dobijeni rezultati i izveštaji

Na osnovu podataka iz Data Warehouse-a sistema, kreirano je 7 izveštaja na osnovu podataka, pri čemu se 2 odnose na prvu temu poslovanja, 2 na drugu temu, i 3 na treću temu.

7.1 Izveštaji za prvu temu poslovanja

Za prvu temu poslovanja koja se odnosi na praćenje priliv novca i broja noćenja napravljeno su dva izveštaja i jedan dodatni.

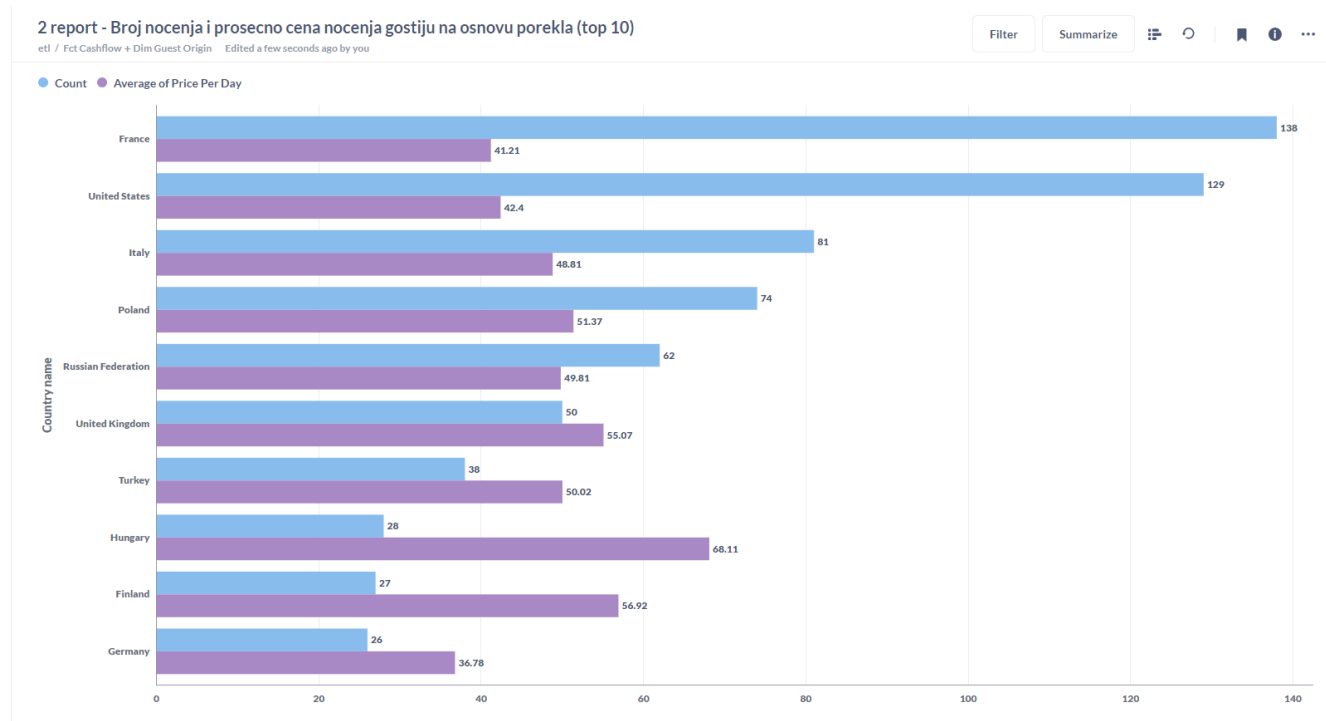
Prvi izveštaj se odnosi na poređenje prihoda koji je svaki od apartmana ostvario u zavisnosti načina oglašavanja za 2022. godinu (Slika 7.1.1). Kako su podaci korišćeni u OLTP šemi (zeleno boja na slici) veštački izgenerisani i ima ih manje razumno je da je i sam prihod ostvaren dosta manji u odnosu na druga dva oglašavača. Može se primetiti da je *Studio with Patio* ostvario skoro 6 puta veći prihod na *Booking* (žuta boja na slici) sajtu u odnosu na *Airbnb* (plava boja na slici) sajt, dok su razlike manje za ostale apartmane. Ova tema omogućava pregled podataka i za druge godine, dok sam izveštaj pruža izdavaču koristan uvid u profite, kao i koji oglašavači postižu uspešnija izdavanja. Poređenje prihoda sa prethodnim godinama može pružiti uvid u moguće postojanje nekih spoljašnjih faktora koji utiču na izdavanje apartmana.



Slika 7.1.2 Prihod po oglašavaču za svaki od apartmana za 2022. godinu

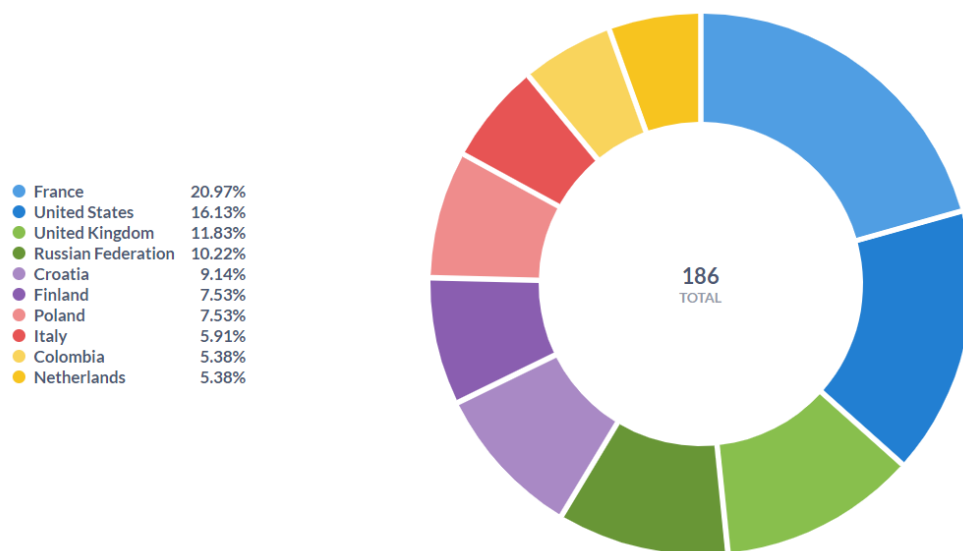
Drugi izveštaj za ovu temu prikazuje ukupan broj noćenja koje su ostvarili gosti različitih porekla (Slika 7.1.2). Plavom bojom je prikazano 10 država čiji posetioци ostvaruju najveći broj noćenja a pored ovoga, ljubičastom bojom, prikazana je i prosečna cena noćenja rezervacije u zavisnosti od države. Može se primetiti da iako osobe koje dolaze iz Mađarske dosta manje rezervišu smeštaje, isto tako su skloniji rezervaciji skupljih smeštaja, ili smeštaja u sezoni kada je cena noćenja viša. Francuzi koji najviše rezervišu apartmane u proseku ih

uzimaju po nižim cenama. Ovo izdavaču pruža uvid koji mu pomaže pri donošenju odluke da li želi više da cilja klijente iz pojedinih država kao i mogućnost primene različitih tehnika reklamiranja, ili jezika za privlačenje drugih gostiju.



Slika 7.1.2 Broj noćenja i prosečna cena noćenja gostiju na osnovu porekla (top 10)

Dodatno je na slici 7.1.3 prikazan i procentualan udeo broja gostiju prema njihovom poreklu u odnosu na ukupan broj nocenja za *Studio with Patio* apartman. Ugostitelju je omogućeno posmatranje ove statistike i za druge apartmane.

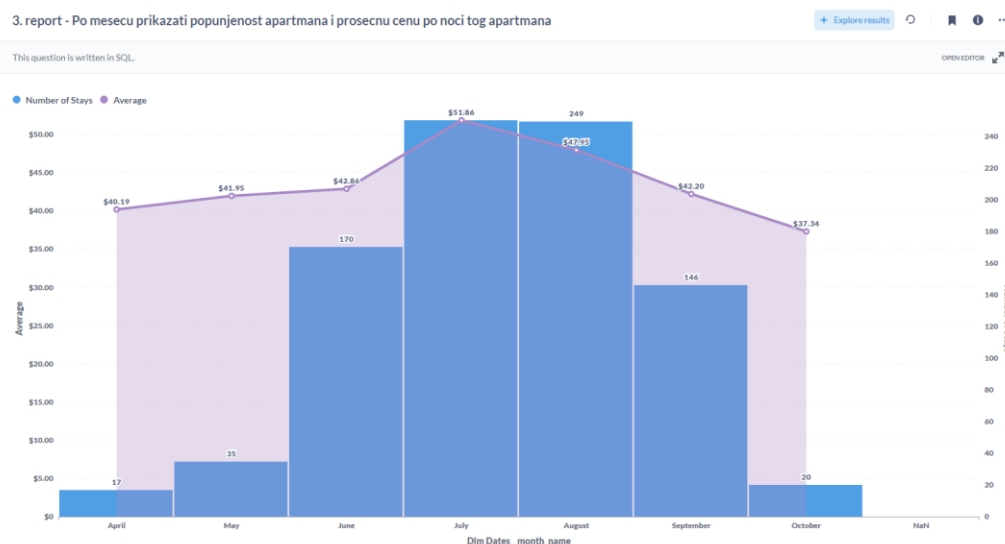


Slika 7.1.3 Udeo broja gostiju prema poreklu za Studio with Patio

7.2 Izveštaji za drugu temu poslovanja

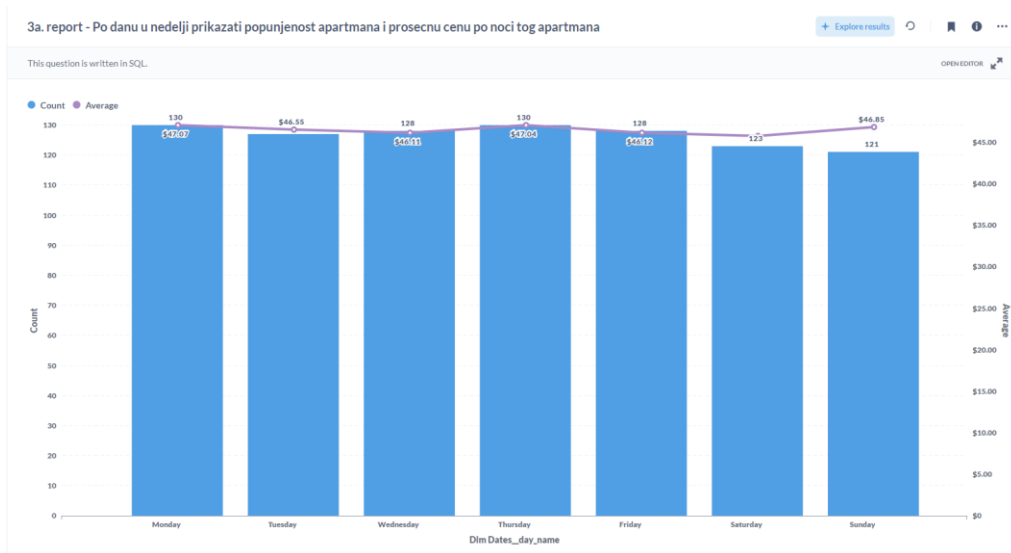
Za drugu temu poslovanja, koja treba da podrži donošenje odluke da li je isplativo da za određeni broj noćenja postoji različita cenovna politika, napravljena su dva izveštaja kao i jedan dodatni.

Prvi izveštaj se odnosi na prikazivanje ukupne popunjenosti smeštajnih kapaciteta, tj. ukupnog broja noćenja svih apartmana za svaki od meseca u godini (Slika 7.2.1). Pored ovoga ljubičastom linijom prikazana je i prosečna cena noćenja po mesecima. Može se uočiti da su smeštaji sezonskog karaktera kako su zatvoreni u periodu od Novembra do Marta meseca. Takođe se može primetiti da u mesecima kada je veća popunjenost i cena noćenja takođe veća. Ovaj izveštaj pomaže izdavaču pri donošenju odluke da u zavisnosti od popunjenosti, ili nepopunjenosti koriguje cene noćenja za određene periode kako bi potencijalno privukao više klijenata.



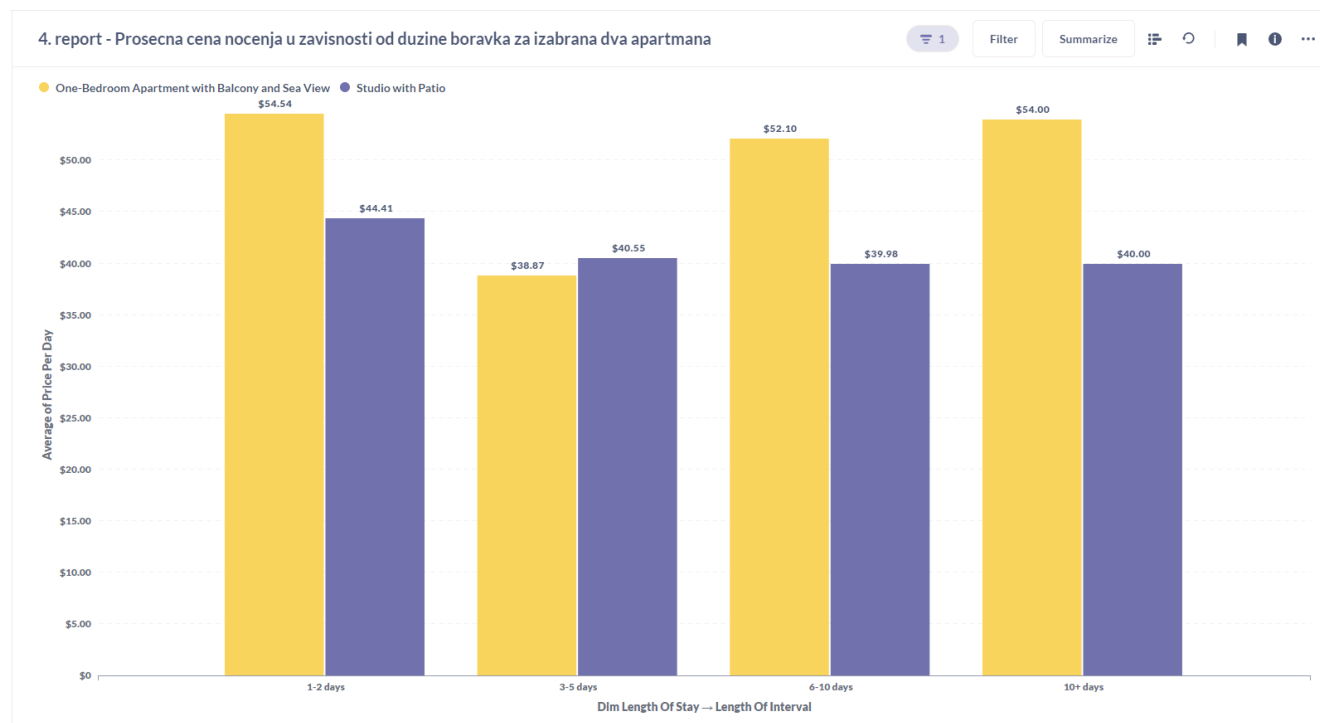
Slika 7.2.1 Popunjenost apartmana i cena noćenja po mesecu

Dodatno su na slici 7.2.1 prikazani broj noćenja i prosečna cena noćenja za svaki od dana u nedelji, kako izdavač bi imao uvid u to da li se pojedinim danima (npr. vikendom) smeštaji češće izdaju, čime bi potencijalno tada povećao cenu. Uočava se da ne postoji velika razlika između broja noćenja tokom različitih dana u nedelji.



Slika 6.2.2 Popunjenost apartmana i cena noćenja po danu u nedelji

Drugi izveštaj za ovu temu prikazuje prosečnu cenu noćenja u zavisnosti od dužine boravka (Slika 7.2.3), pri čemu je dužina boravka data u karakterističnim vremenskim okvirima, 0-2 dana, 3-5 dana, itd. Na slici su prikazane prosečne cene noćea za apartmane *One-Bedroom Apartment with Balcony and Sea View* (žuta boja) i *Studio with patio* (plava boja). Ovde ugostitelj imati priliku da razmotri da li da za duže boravke smanji cenu noćenja kako bi eventualno stimulisao klijente da prave duže rezervacije čime će se postići veća popunjenost smeštaja. Obezbeđeno je donošenje odluke da li želi da ima manju popunjenost po višoj ceni, ili veću popunjenost po manjoj ceni.



Slika 7.2.3 Prosečna cena noćenja u zavisnosti od dužine boravka

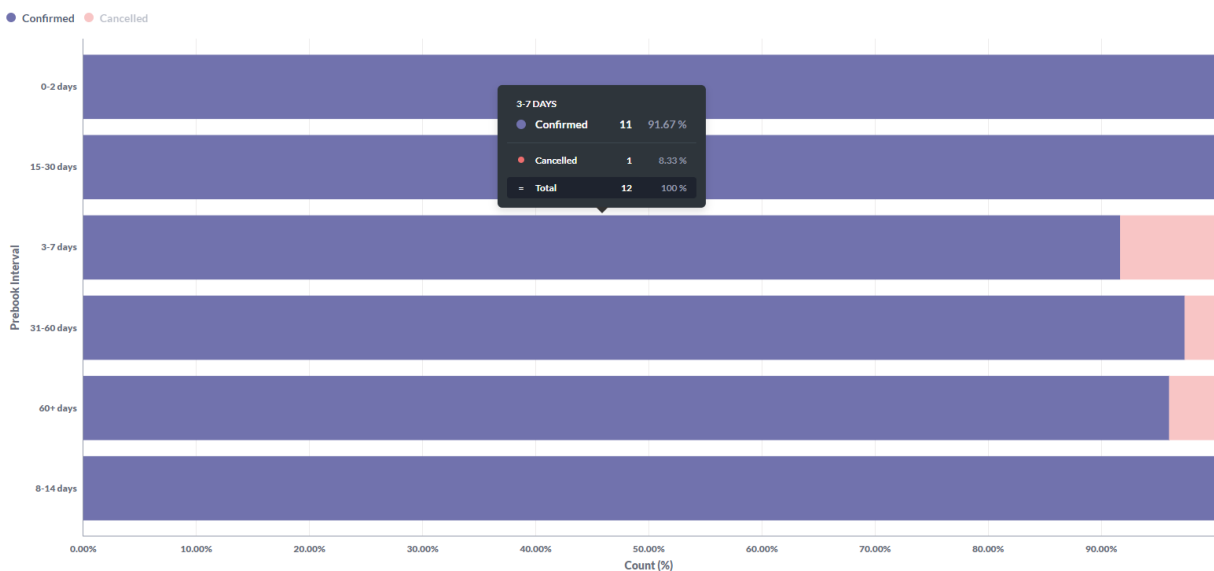
7.3 Izveštaji za treću temu poslovanja

Za poslednju temu poslovanja, koja treba da podrži donošenje odluke na koji način sprovesti otkaznu politiku (*eng. Cancellation policy*), napravljana su tri izveštaja..

Prvi izveštaj se odnosi uporedni prikaz procentualnog odnosa otkazanih i realizovanih rezervacija u odnosu na period koliko je unapred smeštaj rezervisana (Slika 7.3.1), pri čemu su ovi periodi kategorizovani u karakterističnim vremenskim okvirima, 0-2 dana, 7-14 dana, itd. Prelaskom preko bilo kog stuba na izveštaju pruža se uvid i u brojčano stanje otkazanih (roza boja) i neotkazanih (plava boja) rezervacija za dati period. Podaci sa *Airbnb* i *Booking* sajt nisu sadržali otkazane rezervacije, tako da su one veštački napravljene prilikom punjenja inicijalne OLTP šeme podacima. Kako ovih podataka ima malo, samim tim je i broj otkazanih rezervacija takođe mali, pa prikazi na izveštajima nisu toliko reprezentativni kao što bi to bio slučaj u relanoj situaciji. Ovaj izveštaj može pružiti izdavaču koristan prikaz, kao i uviđanje da li su klijenti koji dosta unapred rezervišu smeštaj skloniji njegovom otkazivanju, ili to pak češće rade oni koji kasnije rezervišu smeštaj.

5 report- U zavisnosti od perioda koliko je unapred smestaj rezervisan (ex. 7-14 days) odnos otkazanih i neotkazanih rezervacija i njihovo brojcano stanje za 2022 godinu.

1 Filter Summarize

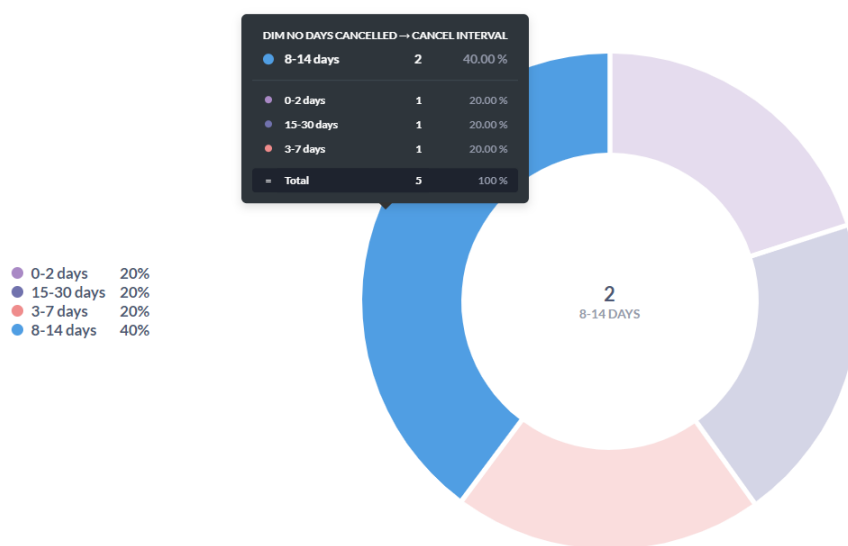


Slika 7.3.1 Odnos otkazanih i neotkazanih rezervacija u zavisnosti od unapred rezervisanog perioda

Drugi izveštaj, koji se može posmatrati kao dodatni izveštaj za ovu temu, prikazuje brojčanu vrednost otkazanih rezervacija u odnosu na ukupan broj otkazanih rezervacija, po periodu koliko je unapred rezervacija otkazana (Slika 7.3.2). Ovo se razlikuje u odnosu na prethodan izveštaj jer je posmatrani period koliko je unapred rezervacija otkazana, za razliku od perioda koliko je unapred rezervacija napravljena. Podaci prikazani na slici su podaci iz 2022. godine. Ovaj izveštaj pruža uvid u to koliko se u odnosu na prvi dan rezervacije najčešće otkazuju rezervacije što može izdavaču poslužiti kao osnova za pomeranje granice kada je potrebno platiti neku kaznu za otkazivanje.

6. report - U zavisnosti od perioda koliko je unapred rezervacija otkazana (ex. 7-14 days) njihova brojcana vrednost, i procenat u odnosu na ukupno otkazanih rezervaija za 2022 godinu.

2



Slika 7.3.2 Procenat otkazanih rezervacija u odnosu na koliko je rezervacija otkazana

Treći, i poslednji izveštaj, ide uz prethodna dva izveštaja i zaokružuje priču potrebnih podataka izdavaču da donese odluku o eventualnoj promeni postojeće cenovne politike. Ovaj izveštaj prikazuje u zavisnosti od perioda koliko je unapred rezervacija napravljena, za otkazanu rezervaciju, prosečna cena dana rezervacije (Slika 7.3.3). Ovde su na slici takođe prikazani podaci za 2022. godinu

Dim Days Prebooked → Prebook Interval	Count	Average of Avg Rate Per Reservation
3-7 days	1	38.29
31-60 days	1	45.7
60+ days	3	44.67

Slika 7.3.3 Prosečna cena rezervacije u odnosu na koliko je unapred rezervisano

8. Zaključak

U ovom dokumentu je prikazano kako se nad sistemom koji se ispituje, putem ECTL procesa prevođenja podataka, došlo do OLAP sistema pogodnog za analizu podataka, kao i samih rezultata ove analize.

Jedan od predloga daljeg unapređenja ovog rada jeste ukрупnjavanje granularnost prve i druge teme koja je na nivou dana, na neki viši nivo, npr. na nivo nedelje. Ovim se postiže smanjenje količine podataka u bazi što i jeste cilj *Data Warehouse* sistema, iako bi se ovim izgubila mogućnost preciznijeg odabira konkretnih vremenskih okvira prilikom zadavanja upita (od prvog tačnog datuma do tačnog drugog datuma). Uz dogovor sa naručiocem projekta moguće je ovo prilagoditi u skladu sa njegovim potrebama.. Takođe, intervalne dimenzije je moguće bolje unaprediti i jasnije odrediti razgovorom sa profesionalcem koji ima iskustva u polju izdavanja apartmana.

Ostali dalji pravci istraživanja i rada na ovo *Data Warehouse* resenju jesu optimizacija i podešavanje nekih parametara fizičke strukture, kao što su indeski, materijalizovani pogledi sa izvedenim i agregiranim podacima, pravljenje OLAP kocki i slično. Razvoj sistema bi se još ogledao i uvođenju automatizovanih operacija ažuriranja, osvežavanja, itd.