



Analiza poslovanja i ponašanja kupaca u lancu kafea

Predlog projekta iz predmeta Sistemi baza podataka

Fakultet tehničkih nauka, Novi Sad

Tim: Tamara Stanković IN 1/2021, Nikola Stojičić IN 11/2021

Datum: 5.9.2025.

O projektu i skupu podataka



Tema: Sveobuhvatna analiza transakcija u lancu kafea sa fokusom na razumevanje kupovnih navika, performansi prodavnica i efikasnosti programa lojalnosti.

Izvor: Kaggle - "G Coffee Shop Transaction Dataset".

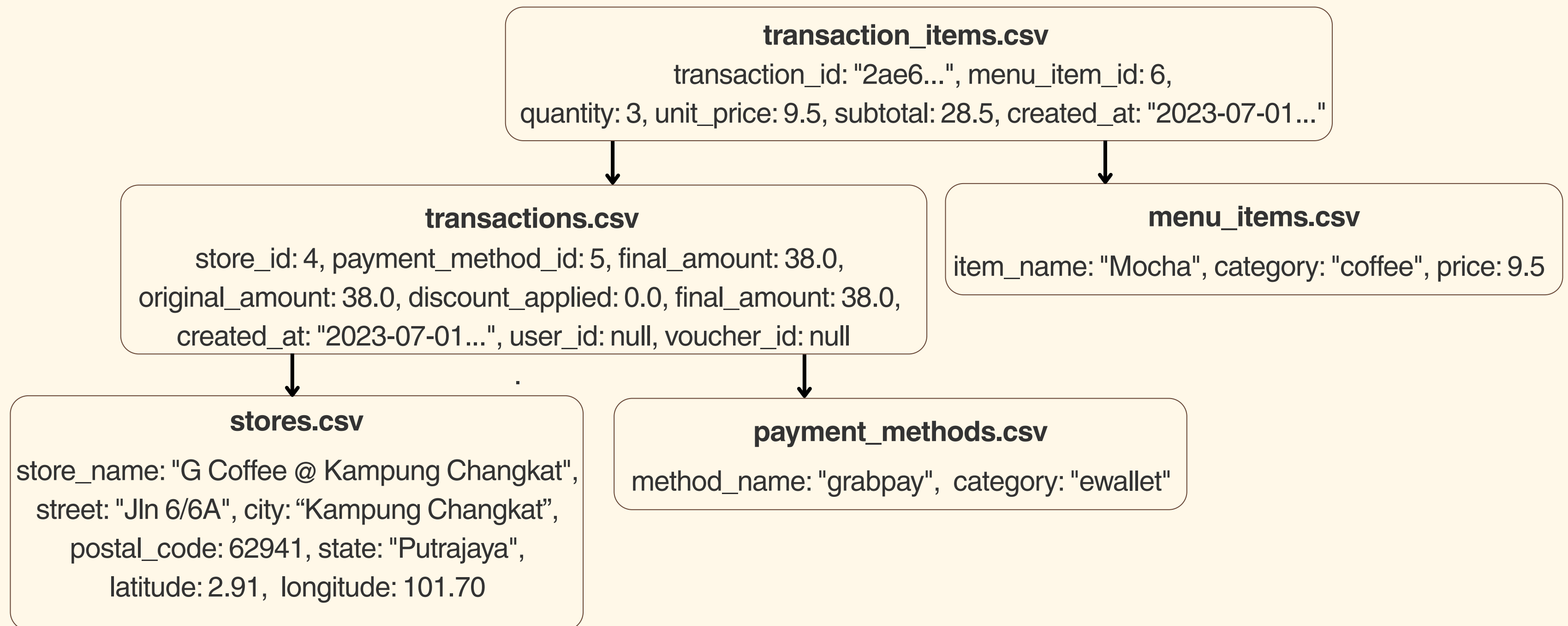
Obuhvat: Analiziramo kompletan set podataka za jul 2023. godine.

Ključne metrike:

- 7 povezanih CSV datoteka
- ~343 MB ukupne veličine
- 620.000+ transakcija
- 1.2 miliona+ prodatih stavki

“Životni put” jedne prodaje kafe

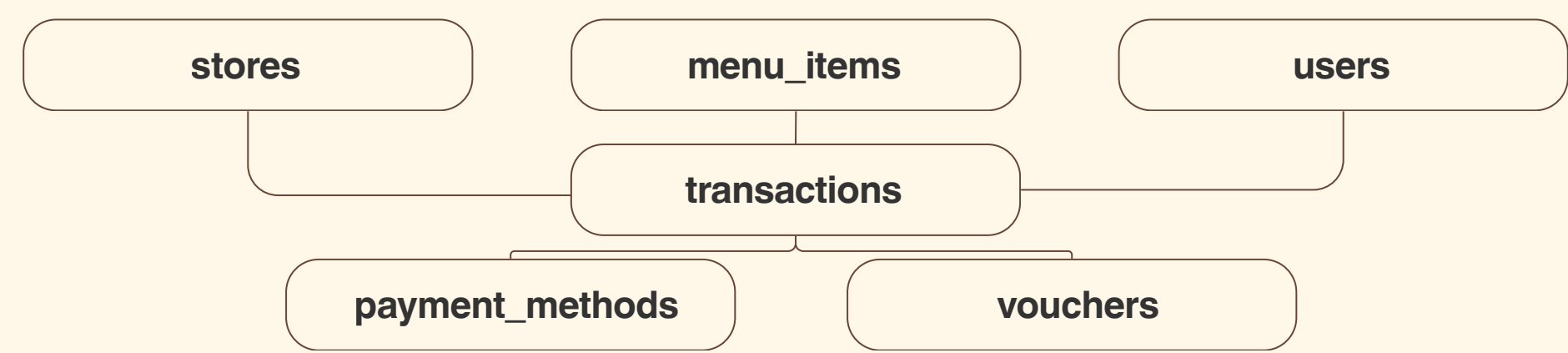
Da bismo demonstrirali kompleksnost, pratimo informacije o jednoj prodatoj kafi, 'Mocha', kroz 5 povezane datoteke:



Zaključak: Kompletna slika se dobija povezivanjem 5 (7) fajlova. Naša MongoDB šema ovo rešava elegantno.

Predlog logičke šeme baze podataka

Naša predložena šema se sastoji od 6 kolekcija. Pet služe kao 'master' tabele sa referentnim podacima, dok je **transactions** centralna kolekcija.



1. Kolekcija: stores

Opis: Svaki dokument predstavlja jednu fizičku lokaciju prodavnice.

Polje	Tip podatka	Semantika
_id	NumberInt	Jedinstveni identifikator prodavnice (store_id).
name	String	Puno ime prodavnice.
address	Object	Ugnježdeni dokument sa adresom.
location	Object	GeoJSON objekat za geoprostorne upite.

address.street - Sttring - Ulica i broj.
address.postal_code - String - Grad u kojem se prodavnica nalazi.
address.state - String - Država ili region.

location.type - String - Point.
location.coordinates - Array - [longitude, lontitude].

2. Kolekcija: menu_items

Opis: Svaki dokument je jedan proizvod iz menija.

Polje	Tip podatka	Semantika
_id	NumberInt	Jedinstveni identifikator proizvoda (item_id).
name	String	Puno ime proizvoda.
category	String	Kategorija proizvoda (npr. "coffee").
price	NumberDecimal	Standardna maloprodajna cena.

Predlog logičke šeme baze podataka (nastavak)

3. Kolekcija: users

Opis: Sadrži podatke o registrovanim korisnicima.

Polje	Tip podatka	Semantika
_id	NumberInt	Jedinstveni identifikator korisnika (user_id).
gender	String	Pol korisnika.
birthdate	Date	Datum rođenja.
registered_at	Date	Datum i vreme registracije.

5. Kolekcija: payment_methods

Opis: "Šifarnik" koji definiše sve dostupne načine plaćanja.

Polje	Tip podatka	Semantika
_id	NumberInt	Jedinstveni identifikator načina plaćanja (method_id).
method_name	String	Naziv načina plaćanja (npr. “credit_card”).
category	String	Šira kategorija načina plaćanja (npr. “card”, “ewallet”).

4. Kolekcija: vouchers

Opis: "Šifarnik" koji sadrži informacije o svim kreiranim vaučerima.

Polje	Tip podatka	Semantika
_id	NumberInt	Jedinstveni identifikator vaučera (voucher_id).
discount_type	String	Tip popusta (npr. "percentage", "fixed_amount").
discount_value	NumberDecimal	Vrednost popusta (npr. 7.0 za 7% ili 5.0 za 5 dolara).
valid_from	Date	Datum od kojeg vaučer važi.
valid_to	Date	Datum do kojeg vaučer važi.

Predlog logičke šeme baze podataka (nastavak)

6. Kolekcija: transation_items

Opis: Sadrži svaku pojedinačnu stavku prodatu u okviru jedne transakcije. Povezuje transakcije sa artiklima sa menija.

Polje	Tip podatka	Semantika
_id	NumberInt	Automatski generisan jedinstveni identifikator stavke.
transaction_id	String	Referenca na _id u transactions kolekciji.
menu_item_id	NumberInt	Referenca na _id u menu_items kolekciji.
unit_price	NumberDecimal	Cena po komadu u trenutku prodaje.
subtotal	NumberDecimal	Ukupan iznos za ovu stavku (quantity * unit_price).

Logička šema - centralna kolekcija transactions

Opis: Najvažnija kolekcija gde svaki dokument predstavlja jedan kompletan račun.

Struktura Dokumenta:

Polje	Tip podatka	Semantika
_id	String	Jedinstveni identifikator transakcije (UUID).
store_id	NumberInt	Referenca na _id u stores kolekciji.
user_id	NumberInt null	Referenca na _id u users kolekciji (null za goste).
payment_method_id	NumberInt	Referenca na _id u payment_methods kolekciji.
voucher_id	NumberInt null	Referenca na _id u vouchers kolekciji.
amounts	Object	Ugnježdeni dokument sa finansijskim podacima.
created_at	Date	Tačan datum i vreme transakcije.

items.menu_item_id - NumberInt - Referenca na _id u menu_items kolekciji.
items.quantity - NumberInt - Količina prodatog artikla za tu stavku.
items.unit_price - NumberDecimal - Cena po komadu u trenutku prodaje.
items.subtotal - NumberDecimal - Ukupan iznos za tu stavku (quantity * unit_price).

amounts.original - NumberDecimal - Iznos pre popusta.
amounts.discount - NumberDecimal - Iznos primenjenog popusta.
amounts.final - NumberDecimal - Konačan iznos računa za naplatu.

Predlog pitanja za agregaciju



Na osnovu preliminarne analize, formulisali smo 10 kompleksnih, poslovno orijentisanih pitanja.

Pitanja zahtevaju napredne tehnike: spajanje više kolekcija, višestruko grupisanje, i kompleksne kalkulacije.

Podeljena su na dva člana tima.

Analitička pitanja - Tamara



1. Analiza "jutarnjeg špica":

Koja su top 3 najprodavanija proizvoda tokom jutarnjeg špica (7-10h) i koliki procenat ukupnog dnevnog prihoda donose?



2. Performanse prodavnica:

Za svaku prodavnicu, izračunati prosečnu vrednost transakcije, ukupan broj transakcija i najprodavaniju kategoriju. Sortirati po prosečnoj vrednosti.



3. Analiza kupovnih navika: registrovani vs. gosti:

Uporediti prosečnu potrošnju i prosečan broj stavki po računu za registrovane korisnike i goste.



4. Popularnost načina plaćanja po vrednosti računa:

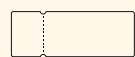
Podeliti transakcije u segmente (mali, srednji, veliki) i za svaki segment prikazati distribuciju načina plaćanja.



5. Analiza lojalnosti korisnika:

Za top 10 najaktivnijih korisnika, izračunati prosečno vreme između njihovih uzastopnih poseta.

Analitička pitanja - Nikola



1. Efektivnost vaučera:

Da li transakcije sa vaučerom imaju veću prosečnu vrednost pre popusta u odnosu na transakcije bez vaučera?



2. Analiza "korpe":

Pronaći 5 najčešćih parova proizvoda koji se kupuju zajedno.



3. Performanse vikendom vs. radnim danima:

Kako se kupovne navike razlikuju između radnih dana i vikenda? Porediti top 3 najprodavanije kategorije proizvoda i njihove udele u prihodu za obe grupe dana.



4. Uticaj kategorija na račun:

Koja kategorija proizvoda najviše doprinosi povećanju vrednosti računa? Za svaku kategoriju, izračunati prosečnu vrednost celokupnog računa na kojem se ta kategorija pojavila.



5. Analiza starosnih grupa korisnika:

Podeliti registrovane korisnike u starosne grupe i za svaku grupu odrediti omiljenu kategoriju proizvoda.

Zaključak

- Odabran je kompleksan i bogat skup podataka koji ispunjava sve zahteve projekta.
- Predložena MongoDB šema je optimizovana za performanse i analizu.
- Definisana su konkretna i izazovna pitanja koja će pokazati dubinsko razumevanje podataka i MongoDB-a.

Hvala na pažnji!

Pitanja?

