

Analiza prodaje namirnica

*Predlog projekta iz predmeta Sistemi
baza podataka*



Opis skupa podataka

Skup podataka sadrži transakcione podatke o prodaji prehrambenih proizvoda u periodu od četiri meseca. Sastoji se od sedam povezanih CSV datoteka koje obuhvataju detalje o prodajama, proizvodima, kupcima, zaposlenima, kao i geografske informacije o gradovima i državama.

Sedam ključnih datoteka:

- sales.csv: Centralna datoteka sa svim prodajnim transakcijama.
- products.csv: Detalji o proizvodima, uključujući cene, klase i kategorije.
- customers.csv: Demografski i geografski podaci o kupcima.
- employees.csv: Informacije o zaposlenima koji su realizovali prodaje.
- categories.csv: Definicije kategorija proizvoda.
- cities.csv & countries.csv: Geografski podaci za segmentaciju tržišta.

Link ka skupu podataka: <https://www.kaggle.com/datasets/andrexbiza/grocery-sales-dataset>

Semantika obeležja

Tabela categories

- **CategoryID:** Jedinstveni identifikator za svaku kategoriju proizvoda.
- **CategoryName:** Naziv kategorije proizvoda.

Tabela cities

- **CityID:** Jedinstveni identifikator za svaki grad.
- **CityName:** Naziv grada.
- **Zipcode:** Poštanski broj.
- **CountryID:** Referenca (strani ključ) na odgovarajuću državu.

Semantika obeležja

Tabela countries

- **CountryID:** Jedinstveni identifikator za svaku državu.
- **CountryName:** Naziv države.
- **CountryCode:** Dvoslovni kod države (npr. RS).

Tabela customers

- **CustomerID:** Jedinstveni identifikator za svakog kupca.
- **FirstName:** Ime kupca.
- **MiddleInitial:** Srednje slovo (inicijal) kupca.
- **LastName:** Prezime kupca.
- **cityID:** Grad kupca (referenca na tabelu 'cities').
- **Address:** Adresa stanovanja kupca.

Semantika obeležja

Tabela employees

- **EmployeeID:** Jedinstveni identifikator za svakog zaposlenog.
- **FirstName:** Ime zaposlenog.
- **MiddleInitial:** Srednje slovo (inicijal) zaposlenog.
- **LastName:** Prezime zaposlenog.
- **BirthDate:** Datum rođenja zaposlenog.
- **Gender:** Pol zaposlenog.
- **CityID:** Jedinstveni identifikator za grad (referenca na tabelu 'cities').
- **HireDate:** Datum zaposlenja.

Tabela products

- **ProductID:** Jedinstveni identifikator za svaki proizvod.
- **ProductName:** Naziv proizvoda.
- **Price:** Cena po jedinici proizvoda.
- **CategoryID:** Jedinstveni identifikator kategorije (referenca).
- **Class:** Klasa proizvoda (Low, Medium, High).
- **ModifyDate:** Datum poslednje izmene.
- **Resistant:** Kategorija otpornosti proizvoda.
- **IsAllergic:** Označava da li je proizvod alergen.
- **VitalityDays:** Klasifikacija vitalnosti proizvoda (npr. rok trajanja u danima).

Semantika obeležja

Tabela sales

- **SalesID:** Jedinstveni identifikator za svaku prodaju.
- **SalesPersonID:** Zaposleni odgovoran za prodaju (referenca).
- **CustomerID:** Kupac koji je izvršio kupovinu (referenca).
- **ProductID:** Prodati proizvod (referenca).
- **Quantity:** Broj prodatih jedinica.

- **Discount:** Primenjeni popust na prodaju.
- **TotalPrice:** Konačna cena prodaje nakon popusta.
- **SalesDate:** Datum i vreme prodaje.
- **TransactionNumber:** Jedinstveni identifikator transakcije.

Predlog logičke šeme

Kolekcija sales

- SalesID
- Quantity
- Discount
- TotalPrice
- SalesDate
- TransactionNumber
- **ProductID** (referenca ka odgovarajućem proizvodu)
- **CustomerID** (referenca ka odgovarajućem kupcu)
- **SalesPersonID** (referenca ka odgovarajućem prodavcu)

Kolekcija products

- ProductID
- ProductName
- Price
- ModifyDate
- Resistant
- Class
- IsAllergic
- VitalityDays
- Category: { CategoryID, CategoryName }

Predlog logičke šeme

Kolekcija customers

- CustomerID
- FirstName
- LastName
- MiddleInitial
- Address
- Location: { CityID, CityName, Zipcode, Country : { CountryID, CountryName, CountryCode } }

Kolekcija employees

- EmployeeID
- FirstName
- LastName
- MiddleInitial
- BirthDate
- Gender
- HireDate
- Location: { CityID, CityName }

PREDLOZI AGREGACIJA

Tijana

- Koji su ukupni prihodi (TotalPrice) i prosečni popusti (Discount) za svaku kategoriju proizvoda po mesecima? Identifikovati najprofitabilnije kategorije i kategorije sa najvećim prosečnim popustom.
- Koji su top 5 najprodavanijih proizvoda (po Quantity i TotalPrice) u celom periodu, i koliki je procenat tih proizvoda označen kao IsAllergic?
- Kako se menjala ukupna prodaja (TotalPrice) i broj transakcija po danima u nedelji tokom celog perioda? Postoje li dani sa značajno većom ili manjom prodajom?
- Postoji li korelacija između Class (klase proizvoda) i Resistant (otpornosti proizvoda) atributa i njihove ukupne prodaje (TotalPrice)? Koji Class/Resistant kombinacije ostvaruju najveći prihod?
- Koje su top 3 kategorije proizvoda koje se najviše prodaju (po TotalPrice) u svakom od top 5 gradova po ukupnoj prodaji?

PREDLOZI AGREGACIJA

Milica

- Koliki je prosečan broj kupovina i prosečna ukupna potrošnja (TotalPrice) po kupcu u različitim gradovima? Identifikovati gradove sa "najlojalnijim" (najviše kupovina/potrošnje) kupcima.
- Koji su top 3 zaposlena (SalesPerson) po ukupnoj prodaji (TotalPrice)? Uporediti njihove demografske podatke (pol, datum zaposlenja) sa prosekom svih zaposlenih.
- Koliko se proizvoda označenih kao IsAllergic prodaje u svakoj zemlji, po Quantity i TotalPrice? Koje zemlje imaju najveću prodaju alergijskih proizvoda?
- Kakav je prosečan 'sadržaj korpe' po različitim državama? Analizirati prosečan broj stavki po transakciji i prosečnu vrednost jedne transakcije za svaku državu.
- Da li gradovi sa većim brojem zaposlenih ostvaruju veću prodaju? Analizirati odnos između ukupnog broja zaposlenih po gradu i ukupnog prihoda ostvarenog od strane kupaca iz istog grada, kako bi se identifikovali gradovi sa najvećom 'produktivnošću' po zaposlenom.

Milica Bošnjak IN₁₈/2021

ČLANOVI TIMA

Tijana Petrović IN₁₀/2021