

SVEUČILIŠTE JURJA DOBRILE U PULI
FAKULTET INFORMATIKE

Tin Pritišanac

Analiza tržišta automobila 1970.-2024.

SEMINARSKI RAD

Pula, rujan, 2025. godine

SVEUČILIŠTE JURJA DOBRILE U PULI
FAKULTET INFORMATIKE

Tin Pritišanac

Analiza tržišta automobila 1970.-2024.

SEMINARSKI RAD

JMBAG: 0171256219, izvanredni student
Studijski smjer: Informatika
Kolegij: Skladišta i rudarenje podataka
Mentor: doc.dr.sc. Goran Oreški

Pula, rujan, 2025. godine



IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisan Tin Pritišanac, ovime izjavljujem da je ovaj seminarski rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio seminarskog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

STUDENT

Pula, rujan, 2025. godine

Sadržaj

| | | |
|----------|---|-----------|
| 1 | Uvod | 1 |
| 2 | Projektni zadatak | 2 |
| 2.1 | Definicija problema i motivacija | 2 |
| 2.2 | Cilj i opis projektnog zadatka | 2 |
| 2.3 | Opseg i ograničenja projekta | 2 |
| 2.4 | Očekivani rezultati i doprinosi | 3 |
| 2.5 | Metodologija rada | 3 |
| 3 | Odabir i analiza skupa podataka | 5 |
| 3.1 | Izvor podataka | 5 |
| 3.2 | Analiza podataka | 5 |
| 3.3 | Priprema podataka | 5 |
| 4 | Relacijski model podataka | 6 |
| 4.1 | Entiteti i atributi | 6 |
| 4.2 | Konceptualni model podataka | 6 |
| 4.3 | Logički model podataka | 6 |
| 4.4 | Kreiranje relacijskog modela podataka | 6 |
| 4.5 | Popunjavanje baze podataka | 6 |
| 4.6 | EER dijagram | 6 |
| 4.7 | Testiranje unosa | 6 |
| 5 | Dimenzijski model podataka | 7 |
| 5.1 | Izrada star scheme | 7 |
| 5.2 | Kreiranje dimenzijskih tablica | 7 |
| 5.3 | Kreiranje tablice činjenica | 7 |
| 6 | ETL proces | 8 |
| 6.1 | Izvlačenje podataka | 8 |
| 6.2 | Transformacija podataka | 8 |
| 6.3 | Popunjavanje skladišta podataka | 8 |
| 7 | OLAP analiza | 9 |
| 7.1 | Definiranje prikaza podataka | 9 |
| 7.2 | Vizualizacija podataka u Tableau | 9 |
| 7.2.1 | Graf 1 TODO | 9 |
| 8 | Zaključak | 10 |
| | Literatura | 11 |
| | Popis slika | 12 |
| | Popis tablica | 13 |

1 Uvod

U današnjem digitalno vođenom poslovnom okruženju, sposobnost efikasnog prikupljanja, obrade i analize velikih količina podataka postala je ključni čimbenik uspjeha za organizacije u gotovo svim industrijskim granama. Automobilska industrija, kao jedna od najkompleksnijih i najkonkurentnijih grana gospodarstva, posebno se oslanja na napredne tehnologije skladištenja i analize podataka kako bi razumjela tržišne trendove, potrebe kupaca i operacijske učinkovitosti [1].

Skladišta podataka (eng. *data warehouses*) predstavljaju temelj modernih sustava za podršku odlučivanju, omogućujući integraciju različitih izvora podataka u jedinstvenu, koherentnu strukturu optimiziranu za analitičke potrebe [2]. U kontekstu automobilske industrije, ovakvi sustavi omogućuju analizu složenih odnosa između cijena vozila, karakteristika proizvođača, tržišnih segmenata i regionalnih specifičnosti.

Ovaj rad predstavlja sveobuhvatan pristup razvoju sustava za skladištenje i rudarenje podataka primjenjenog na analizu automobilske tržišta. Kroz razvoj kompletnog ETL (Extract, Transform, Load) procesa, projekt demonstrira transformaciju sirovih podataka o automobilima u strukturirani dimenzijski model prilagođen OLAP (Online Analytical Processing) analizama. Korištenjem skupa podataka koji sadrži preko 97.000 zapisa o vozilima različitih proizvođača, modela i karakteristika, razvijen je sustav koji omogućuje dubinsku analizu tržišnih trendova i poslovnih uvida.

Glavni cilj ovog projekta je ilustracija praktične primjene teorijskih koncepata skladišta podataka kroz razvoj funkcionalnog sustava koji može poslužiti organizacijama poput autoklubova, analitičkih kuća ili samim proizvođačima automobila u donošenju informiranih poslovnih odluka [3]. Projekt obuhvaća sve ključne faze razvoja sustava - od eksploratorne analize početnih podataka, preko dizajna relacijskog i dimenzijskog modela, do implementacije ETL procesa i prijedloga OLAP analiza.

Struktura rada prati logični tijek razvoja sustava, počevši od analize i pripreme početnog skupa podataka, preko stvaranja normaliziranog relacijskog modela, do konačne implementacije zvjezdastog modela optimiziranog za analitičke potrebe. Svaki korak popraćen je detaljnim objašnjenjima projektnih odluka i praktičnih implementacijskih izazova, čineći ovaj rad korisnim resursom za razumijevanje kompleksnosti razvoja realnih sustava za skladištenje podataka.

Kroz ovaj projekt, nastoji se pokazati kako tehnologije poput Apache Spark-a, MySQL-a i Tableau-a mogu biti integrirane u koherentan sustav koji omogućuje ne samo tehnički ispravan rad, već i stvaranje dodane vrijednosti kroz kvalitetne poslovne uvide [4].

2 Projektni zadatak

2.1 Definicija problema i motivacija

Automobilska industrija je jedna od najkompleksnijih gospodarskih grana koja generira ogromne količine podataka - od osnovnih karakteristika vozila, preko cijena i tržišnih trendova, do regionalnih specifičnosti i preferencija kupaca. Organizacije koje se bave analizom automobilskeg tržišta, poput autoklubova, analitičkih kuća ili samih proizvođača, suočavaju se s izazovom pretvaranja ovih velikih količina podataka u korisne poslovne uvide.

Tradicionalni pristup analize podataka korištenjem jednostavnih baza podataka i osnovnih alata za izvještavanje često se pokazuje neadekvatnim za složene analitičke potrebe. Potreba za dubinskim analizama tržišnih trendova, usporednim analizama proizvođača, segmentacijom kupaca i predviđanjem buduće dinamike tržišta zahtijeva sofisticiraniji pristup [1].

Upravo tu se prepoznaje potreba za razvojem naprednog sustava skladištenja i analize podataka koji će omogućiti organizacijama da iz sirovih podataka izvuku maksimalnu vrijednost te donose informirane strateške odluke.

2.2 Cilj i opis projektnog zadatka

Glavni cilj ovog projekta je razvoj potpunog sustava za skladištenje i rudarenje podataka prilagođenog analizi automobilskeg tržišta. Projekt obuhvaća kompletan tijek rada - od početnih sirovih podataka do konačnih analitičkih izvještaja koji mogu koristiti stvarnim organizacijama u njihovom poslovanju.

Konkretno, projekt ima za cilj:

- **Dizajnirati i implementirati relacijski model podataka** koji odgovara strukturi realnih podataka o automobilima, uključujući sve važne entitete i njihove međusobne odnose
- **Razviti dimenzijski model (star schema)** optimiziran za OLAP analize, koji omogućuje efikasno izvršavanje složenih analitičkih upita
- **Implementirati ETL proces** koji transformira podatke iz relacijske strukture u dimenzijski model koristeći suvremene tehnologije poput Apache Spark-a
- **Kreirati sustav za OLAP analize** koji demonstrira praktičnu primjenu različitih analitičkih operacija (slice, dice, drill-down, roll-up, pivot)
- **Pokazati poslovnu vrijednost** kroz konkretne scenarije korištenja koji ilustriraju kako ovakav sustav može koristiti organizacijama u donošenju poslovnih odluka

2.3 Opseg i ograničenja projekta

Projekt se fokusira na analizu podataka o automobilima prodavanima na fiktivnom tržištu, koristeći skup od preko 97.000 zapisa koji obuhvaća vozila različitih

proizvođača, modela i karakteristika. Ovaj skup podataka pruža reprezentativan uzorak koji omogućuje demonstraciju svih ključnih koncepata skladišta podataka.

Vremenski okvir podataka pokriva razdoblje od 1970. do 2024. godine, s naglaskom na zadnja dva desetljeća, što omogućuje analizu dugoročnih trendova i promjena na tržištu. Podaci uključuju ključne attribute poput cijene, godine proizvodnje, kilometraže, vrste goriva, veličine motora i drugih karakteristika relevantnih za tržišnu analizu.

Projekt se ograničava na demonstraciju tehničkih mogućnosti i metodologija, a ne pretendira na potpunu komercijalnu implementaciju. Fokus je na edukacijskim aspektima i ilustraciji najboljih praksi u razvoju sustava za skladištenje podataka.

2.4 Očekivani rezultati i doprinosi

Na završetku projekta očekuje se sljedeće:

1. **Funkcionalno skladište podataka** s implementiranim relacijskim i dimenzijskim modelom podataka
2. **Potpuno funkcionalan ETL proces** koji automatizira transformaciju podataka između različitih modela
3. **Demonstracija OLAP mogućnosti** kroz konkretne analitičke scenarije i upite
4. **Dokumentacija procesa** koja može služiti kao vodič za buduće slične projekte
5. **Praktični uvidi u tržište automobila** koji ilustriraju poslovnu vrijednost analitičkih sustava

Ovaj projekt predstavlja praktičnu demonstraciju kako teorijski koncepti skladišta podataka i business intelligence mogu biti primijenjeni u realnom kontekstu, pružajući studentima i praktičarima vrijedan uvid u izazove i mogućnosti moderne analize podataka [2].

2.5 Metodologija rada

Projekt slijedi strukturiran pristup razvoja sustava za skladištenje podataka, koji se može podijeliti u pet glavnih faza:

1. **Eksploratorna analiza podataka** - detaljno istraživanje početnog skupa podataka radi razumijevanja strukture, kvalitete i potencijalnih izazova
2. **Dizajn relacijskog modela** - stvaranje normalizirane strukture podataka koja odražava realne poslovne entitete i njihove odnose
3. **Implementacija dimenzijskog modela** - razvoj star schema arhitekture optimizirane za analitičke potrebe

4. **ETL proces** - implementacija sustava za ekstrahiranje, transformaciju i učitavanje podataka koristeći Apache Spark
5. **OLAP analize** - demonstracija analitičkih mogućnosti kroz praktične scenarije kreiranja grafova i tablica u programu Tableau

Svaka faza dokumentirana je s objašnjenjima projektnih odluka, tehničkih izazova i načina njihova rješavanja, čineći projekt korisnim resursom za razumijevanje praktičnih aspekata razvoja sustava za skladištenje podataka.

3 Odabir i analiza skupa podataka

3.1 Izvor podataka

3.2 Analiza podataka

3.3 Priprema podataka

4 Relacijski model podataka

4.1 Entiteti i atributi

4.2 Konceptualni model podataka

4.3 Logički model podataka

4.4 Kreiranje relacijskog modela podataka

4.5 Popunjavanje baze podataka

4.6 EER dijagram

4.7 Testiranje unosa

5 Dimenzijski model podataka

5.1 Izrada star scheme

5.2 Kreiranje dimenzijskih tablica

5.3 Kreiranje tablice činjenica

6 ETL proces

6.1 Izvlačenje podataka

6.2 Transformacija podataka

6.3 Popunjavanje skladišta podataka

7 OLAP analiza

7.1 Definiranje prikaza podataka

7.2 Vizualizacija podataka u Tableau

7.2.1 Graf 1 TODO

8 Zaključak

Literatura

- [1] N. Silva, J. Barros, M. Y. Santos, C. Costa, and P. Cortez. Advancing Logistics 4.0 with the Implementation of a Big Data Warehouse: A Demonstration Case for the Automotive Industry. *Electronics*, 10(18):2221, 2021. [Na internetu]. Dostupno: <https://www.mdpi.com/2079-9292/10/18/2221> [pristupano 29. kolovoza 2025.].
- [2] G. Garani, A. Chernov, and I. Savvas. A Data Warehouse Approach for Business Intelligence. In *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pages 70–75, 2019. [Na internetu]. Dostupno: <https://ieeexplore.ieee.org/document/8795395> [pristupano 29. kolovoza 2025.].
- [3] P. Nima. Data Warehousing and Business Intelligence Project on Car Insights in the United Kingdom. ResearchGate, 2018. [Na internetu]. Dostupno: <https://www.researchgate.net/publication/330837638> [pristupano 29. kolovoza 2025.].
- [4] B. Leka, D. Leka, and B. Baraku. Driving Operational Excellence: Business Intelligence in the Car Parts Industry. *WSEAS Transactions on Business and Economics*, 22:356–367, 2025. [Na internetu]. Dostupno: <https://www.wseas.com/journals/bae/2025/a505118-356.pdf> [pristupano 29. kolovoza 2025.].

Popis slika

Popis tablica