Analiza podataka primjenom MS Excel-A I SQL alata

Protega, Vinka

Undergraduate thesis / Završni rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Polytechnic of Sibenik / Veleučilište u Šibeniku**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:143:811388

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-04-19**

Repository / Repozitorij:

VUS REPOSITORY - Repozitorij završnih radova Veleučilišta u Šibeniku

VELEUČILIŠTE U ŠIBENIKU
ODJEL POSLOVNE INFORMATIKE
PREDDIPLOMSKI STRUČNI STUDIJ
POSLOVNA INFORMATIKA

Vinka Protega

ANALIZA PODATAKA PRIMJENOM MS EXCEL-A I SQL ALATA

Završni rad

Šibenik, 2022.

VELEUČILIŠTE U ŠIBENIKU
ODJEL POSLOVNE INFORMATIKE
PREDDIPLOMSKI STRUČNI STUDIJ
POSLOVNA INFORMATIKA

ANALIZA PODATAKA PRIMJENOM MS EXCEL-A I SQL ALATA

Završni rad

Kolegij: Poslovni informacijski sustavi
Mentor: dipl.ing. Želimir Mikulić viši predavač
Studentica: Vinka Protega
Matični broj studentice: 1219001144

Šibenik, rujan,2022.

Ja, Vinka Protega, studentica Veleučilišta u Šibeniku, JMBAG 1219001144 izjavljujem pod materijalnom i kaznenom odgovornošću i svojim potpisom potvrđujem da je moj završni rad na preddiplomskom stručnom studiju Poslovna informatika pod naslovom:

Analiza podataka primjenom MS Excel-a i SQL alata

isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju korištene bilješke i bibliografija.
Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava.
Izjavljujem, također, da nijedan dio rada nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

U Šibeniku, 15.rujna 2022.

Studentica:Vinka Protega

Vinka Protega

TEMELJNA DOKUMENTACIJSKA KARTICA

Veleučilište u Šibeniku Završni rad

Odjel Poslovne Informatike

Preddiplomski stručni studij Poslovna informatika

ANALIZA PODATAKA PRIMJENOM MS EXCEL-A I SQL ALATA

VINKA PROTEGA

Ulica branitelja domovinskog rata 7,Šibenik , vprotega@vus.hr

Rad donosi kratki uvid u dva Microsoft alata za analizu podataka, Excel i Microsoft SQL Server. Funkcionalnosti svakog od alata predstavljene su u teoretskom i praktičnom dijelu rada. U radu je korišten isti set podataka za prikaz praktičnog dijela alata. U prvom dijelu predstavljen je Excel kroz njegova najznačajnija svojstva kad je riječ o analizi podataka, a to su built-in formule, pivot tablice i pivot grafikoni. U ovo dijelu predstavljena je i nova funkcionalnost programa Excel, Power Query. Drugi dio odnosi se na Mirosoft SQL Server, kao nezaobilazan alat za analizu podataka prikupljenih u bazama podataka. Ovaj dio donosi kratki pregled razvoja SQL, vrste naredbi i vrste funkcija. BI i skladišta podataka opisana su prije praktičnog dijela rada. Predstavljene su glavne razlike OLTP sustava i BI sustava te organizacijske sheme skladišta podataka. U praktičnom dijelu izrađena je baza podataka, a pripadajuće tablice napunjene su korištenjem još jednog Microsoft alata SQLServer 2019 Import Export Data. U praktičnom dijelu prikazani su upiti na bazu podataka i detaljno objašnjeni dobiveni rezultati. Rad završava zaključkom o prednostima i nedostacima navedenih alata.

(36 stranica/ 24 slike / 4 tablice / jezik izvornika: hrvatski)

Rad je pohranjen u digitalnom repozitoriju Knjižnice Veleučilišta u Šibeniku

Ključne riječi: *analiza podataka, SQL, Excel, pivot tablice*

Mentor: dip.ing Želimir Mikulić viši predavač

Rad je prihvaćen za obranu dana:

BASIC DOCUMENTATION CARD

Polytechnic of Šibenik
Department of Business Informatics
Professional Undergraduate Studies of Business Informatics

Bachelor Thesis

Data Analitics Using MS Excel and SQL Tools

VINKA PROTEGA
Ulica branitelja domovinskog rata 7
vprotega@vus.hr

Abstract:
This paper provides a brief insight into two Microsoft tools for data analysis, Excel and Microsoft SQL Server. The functionalities of each of the tools are presented in the theoretical and practical parts of the paper. The paper uses the same data set to display the practical use of the tool. In the first part of the paper, Excel is presented through its most important features when it comes to data analysis, namely built-in formulas, pivot tables and pivot charts. In this part, the new functionality of the Excel program, Power Query, is presented. The second part refers to Microsoft SQL Server, as one of the most important tools for analyzing data stored in databases. This section provides a brief overview of SQL development, statement types, and function types. BI and data warehouses are described before the practical part. The main differences between the OLTP system and the BI system and the organizational chart of the data warehouse are presented. In the practical part, a database was created, and the associated tables were filled using another Microsoft tool, SQLServer 2019 Import Export Data. Data analysis has been done through queries and the obtained results are explained in detail. The paper ends with a conclusion on the advantages and disadvantages of the mentioned tools.

(36 pages / 24 figures / 4 tables / original in Croatian language)
Thesis deposited in Polytechnic of Šibenik Library digital repository
Keywords: *data analysis, SQL, Excel, pivot tables*
Supervisor: Želimir Mikulić, senior lecturer

Paper accepted:

Sadržaj

1. UVOD 1
2. METODE I TEHNIKE RADA 2
3. ANALIZA PODATAKA 3
3.1 Razvoj analize podataka 3
3.2 Faze analize podataka 4
4. ALATI ZA ANALIZU PODATAKA 5
4.1 Excel 5
4.1.1 Power Query 5
4.1.1.1 Praktični primjer korištenja Power Query mehanizma 8
4.1.2 Analiza podataka korištenjem built in funkcija 12
4.1.2.1 Praktična primjer primjene analize podataka primjenom Excel funkcija 14
4.1.3 Analiza podataka korištenjem zaokretnih tablica i grafikona 17
4.1.3.1 Praktični primjer primjene analize podataka primjenom Pivot tablica 18
i grafikona 18
4.2 Microsoft SQL Server Management Studio 21
4.2.1 SQL -Structured Query Language 22
4.2.2 Funkcije u T-SQL 24
4.2.3. Online obrada transakcija (OLTP) vs BI (Business Intelligence) 25
4.2.3.1. Analiza podataka u SSMS – praktičan primjer 28
5. ZAKLJUČAK 36
LITERATURA 37
PRILOZI 37
1.1 Popis slika 37
1.2 Popis tablica 38

1. UVOD

Razvoj tehnologije, mogućnost stvaranja, pohranjivanja i analiziranja ogromne količine podataka te poslovni ekosustav koje se strelovito razvija i mijenja, stavili su analizu podataka u centar pozornosti. Dobiti kvalitetan, komercijalno relevantan uvid u vlastite i/ili iz vanjskih izvora prikupljene podatke može biti ključno u otkrivanju novih tržišta, tržišnih niša ili stvaranju prednosti pred konkurentskim organizacijama.

Jasno da ovaj koncept nije ništa novo. Poslovni svijet odavno koristi uvide u podatke iz prošlosti poslovanja, da bi anticipirali buduće događaje, kreirali nove strategije itd. Ono što je danas drugačije je definicija što su uopće podaci i tehnološki napredak koji nam omogućava da ih pohranjujemo, analiziramo i iz njih ekstrahiramo informacije obradom takvog volumena podataka koji je u, ne tako davnoj, prošlosti bio gotovo nezamisliv.

Valjanost i točnost dobivenih informacija uvelike ovisi o tome koliko su jasno definirana strateška pitanja na koja želimo odgovor i koliko su kvalitetni podaci koje koristimo.

Analiza podataka nezamjenjiva je u otkrivanju skrivenih obrazaca, korelacija i drugih uvida, dohvatom i analiziranjem velikih količina podataka u relativno kratkom vremenu. Podatkovni analitičari redovito se susreću s promjenama, novim zahtjevima i algoritmima u analizi podataka, to je jedan od razloga zašto mnogi novi alati ulaze na tržište i zašto drugi alati gube na važnosti.

Odabir odgovarajućeg alata izuzetno je važan jer krivo izvedena ili neodgovarajuća analitika može imati dalekosežne posljedice na kvalitetu odluka koje se temeljem njih donose a time i cjelokupno poslovanje pa u konačnici i opstanak organizacije. Zato je važno pri odabiru odgovarajućih alata imati jasno definiranu misiju i viziju organizacije, kako će se novi alat uklopiti u poslovne ciljeve, tko su korisnici (tko će se koristiti dobivenim uvidima i izvješćima), analizirati postojeći informacijski sustav radi stjecanja uvide u način, veličinu ili troškove migracije na novo okruženje i ne manje važno, troškove nabavke novog sustava prilagoditi planiranom rastu organizacije.

Cilj analize podataka je dobiti uvid u prikupljene podatke, najčešće u obliku izvještaja, stoga izvještavanje treba prilagoditi korisnicima kojima je namijenjeno. Dobra komunikacija s korisnicima preduvjet je efektivne i efikasne analize podataka. U ovom radu prikazana je analiza jednog većeg seta podataka kroz različite alate od kojih neki uključuju vizualizaciju, a neki od alata ograničeni su na fundamentalno izvještavanje u obliku različitih tablica. Za postizanje optimalnih rezultata kombiniranje različitih alata za analizu podataka pokazalo se kao najbolji izbor.

2. METODE I TEHNIKE RADA

Za pisanje ovog rada bilo je potrebno pripremiti i analizirati relevantnu literaturu. U radu su korišteni različiti izvori kao što se članci, web stranice, online tečajevi te različiti internetski izvori čiji se detaljan popis nalazi na kraju ovog rada. Set podataka koji služi kao podloga za prikaz praktičnog dijela analize podataka preuzet je s internetske stranice www.simplilearn.com, sve ostale radnje i manipulacija podacima izrađene su od strane autora isključivo za potrebe ovog rada.

Kako se razvijalo područje analize podataka sukladno tome rastao je i raspon alata dostupnih za analizu. Ovaj rad donosi kratki uvid u jedan dio Microsoft alata za analizu podataka, a to su Excel sa svojim dodacima („add-in“) i Microsoft SQL Server Management Studio. Ovi alati su odabrani jer su dostupni i besplatni učenicima i studentima u RH, a istodobno su relevantni i nezaobilazni u svakoj ozbiljnijoj analizi podataka.

Excel , kao jedan od najčešće korištenih alata za analizu podataka , zauzima i najveći dio rada. U potpoglavljima opisane su neke (ali ne i sve) funkcionalnosti Excela kao što su built-in funkcije, Power Query, Pivot (zaokretne) tablice te pivot grafikoni koje opravdavaju njegovu 30-to godišnju dominaciju ovim područjem.

Rad se sastoji od teoretskog dijela koji donosi kratki prikaz različitih funkcionalnosti i mogućnosti svakog od navedenih alata i praktičnog dijela u kojem je prikazana analiza podataka. Isti set podataka prilagođen je i korišten u sva tri alata jer tako najbolje dolaze do izražaja mogućnosti, prednosti i nedostaci pojedinog alata.

3. ANALIZA PODATAKA

Jedna od definicija analize podataka opisuje taj pojam kao:
„ tehnika kojom se iz kvantitativnih podataka dobivaju novi kvantitativni pokazatelji i osnova za kvalitativne ocjene. Iz različitih statistika ili istraživanja dodatnim obradbama podataka i analizama može se dobiti nova slika pojave. U potonjem slučaju radi se o »sekundarnoj analizi«.[1]
Jednostavnijim rječnikom rečeno analiza podataka je proces čišćenja, transformacije i analize sirovih podataka kako bi se dobile korisne, relevantne informacije koje mogu pomoći organizacijama u donošenju pametnih odluka.

3.1 Razvoj analize podataka

U posljednjih nekoliko godina analiza podataka postala je izuzetno popularna u poslovnom svijetu zahvaljujući razvoju tehnologije . Ipak ovo i nije tako nov pojam. Prva upotreba podataka u svrhe poslovne analitike datira još iz 1865. Tijekom tog vremena, sir Henry Furnese, bankar, aktivno prikupljajući informacije i djelujući u skladu s njima, uvijek je bio korak ispred svojih konkurenata. To jasno govori da su se i u prošlosti menadžeri tvrtki, poput Sir Furnesea, više oslanjali na podatke i empirijske dokaze, a nego na instinkte.
Ono što danas podrazumijevamo pod pojmom analiza podataka, svoj ubrzani razvoj duguje isključivo razvoju računala.
1956. godine, tehnološki div, IBM je izumio prvi pogon tvrdog diska. Daljnji razvoj osobnih računala omogućilo je korisnicima prikupljanje velike količine podataka uz veću fleksibilnost. 90-tih godine prošlog stoljeća ostvaruje se značajan napredak na polju prikupljanja i čuvanja podataka. Do tada je bila dostupna značajna količina podataka, ali nije bilo centraliziranog mjesta za njihovu pohranu. Rješenje problema ponudili su Ralph Kimball i Bill Inmon u vidu izgradnje skladišta podataka (Data Warehouse).
Kimball (1996)definira skladište podataka kao kopiju transakcijskih podataka specijalno strukturirana za upite i analize.
Novi milenij donosi različita analitička rješenja u obliku poslovne inteligencije.[2]

[1] analiza podataka. *Hrvatska enciklopedija, mrežno izdanje.* Leksikografski zavod Miroslav Krleža, 2021. Pristupljeno 31. 8. 2022. <http://www.enciklopedija.hr/Natuknica.aspx?ID=2466>.
[2] IT i konzultantska tvrtka Gartner skovala je riječ 'poslovna inteligencija (BI)' za „sveobuhvatan koncept koji uključuje tehnologije, tehniku , instrumente i najbolje prakse koji omogućavaju stjecanje i vrednovanje znanja kako bi se poboljšalo i ojačalo donošenje odluka

Tvrtke koje već tada prednjače u njihovoj izradi su Oracle, SAP, IBM…
Početkom 2005. tvrtke su počele usmjeravati svoje napore na poboljšanje brzine kojom bi informacije bile dostupne. Uvedeni su novi alati za analizu podataka kako bi se osiguralo prikupljanje (rudarenje) podataka i uvid u prikupljene podatke tehničkom i ne-tehničkom osoblju.
Otprilike u to vrijeme, sve veća povezanost poslovnog svijeta dovela je do potrebe za informacijama u stvarnom vremenu. Budući da je internet dostupan gotovo svima i stvara sve više podataka, tvrtke su trebale bolja rješenja za pohranu i analizu svih podataka. Izgradnja računala s većim kapacitetom za pohranu i boljom brzinom mnogima nije bila moguća, pa su tvrtke pribjegavati korištenju nekoliko strojeva istovremeno. Ovo je bio početak računalstva u oblaku. Sukladno rastu količine podataka, razvoju načina skladištenja i brzini kojom izvještavanje postaje dostupno, razvijali su se i alati za analizu podataka.

3.2 Faze analize podataka

Analiza podataka je proces čišćenja, transformacije i analize sirovih podataka kako bi se dobile korisne, relevantne informacije koje mogu pomoći tvrtkama u donošenju pametnih odluka. Dajući relevantne uvide u podatke, koji se obično prikazuju u tablicama i grafikonima, analiza podataka pomaže u smanjenju rizika povezanih s donošenjem odluka.
Analitika podataka ne obuhvaća samo analizu, već i prikupljanje, organizaciju, pohranu podataka te alate i tehnike koje se koriste za dublje uvide u podatke, kao i one tehnike koji se koriste za predstavljanje rezultata, kao što su alati za vizualizaciju podataka.
S druge strane, analiza podataka se bavi procesom transformacije sirovih podataka u smislene statistike, informacije i objašnjenja.
Iterativni proces analize podataka sastoji se od sljedećih faza:

1. Definiranje zahtjeva za podacima
2. Prikupljanje podataka
3. Obrada podataka
4. Čišćenje podataka
5. Analiza podataka
6. Vizualizacija podataka

4. ALATI ZA ANALIZU PODATAKA

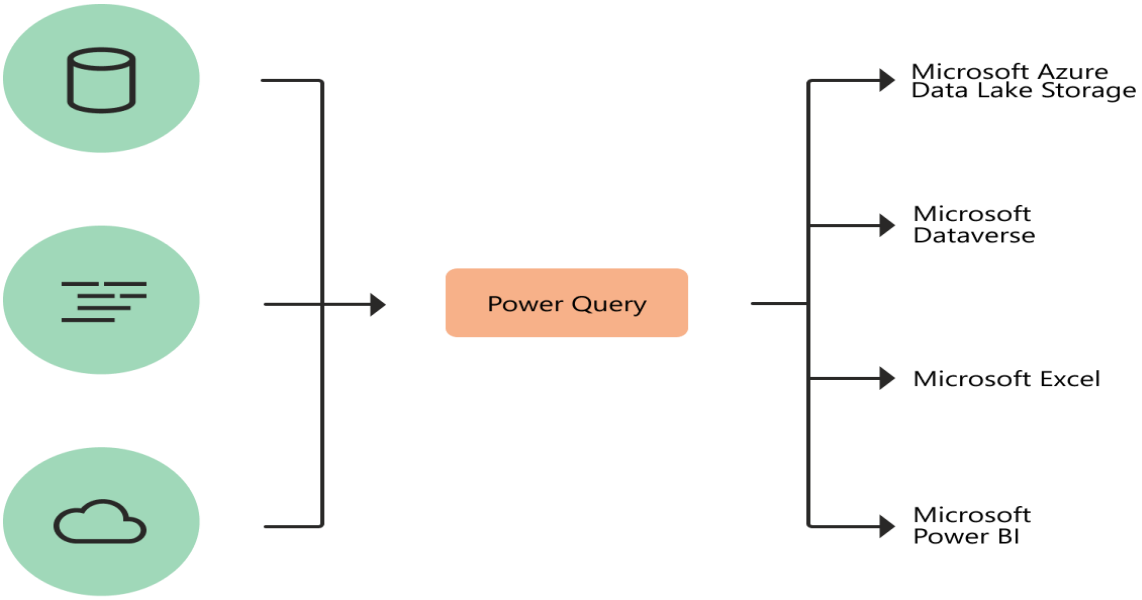
4.1 Excel

Microsoft Excel jedan je od najčešće korištenih programa za analizu podataka, a ugrađene zaokretne (pivot) tablice najpopularniji su analitički alat. Jedan od razloga popularnosti Excela je svakako činjenica da je relativno jeftin, dostupan i ne osobito kompliciran alat s kojim je, barem donekle, upoznata većina ljudi što značajno smanjuje troškove dodatne edukacije zaposlenika, a svojim performansama čini odličnu podlogu za prikupljanje, analizu i vizualizaciju podataka.
Korisniku je dostupan i čitav niz pod-alata i mehanizama za oblikovanje, analizu i vizualizaciju podataka kao što su uvjetno oblikovanje, tablice, različite funkcije, analiza što ako, rješavači(Solver), model podataka, PowerPivot, PowerView, Power Map i druge naredbe programa Excel.

4.1.1 Power Query

Kao što ime sugerira, Power Query (PQ) je vrlo moćan mehanizam u Excelu za automatizaciju dohvaćanja podataka. PQ omogućuje korisniku kompletan ETL (extract, transform, load) proces. Uvoz, odnosno ekstrakciju podataka iz različitih izvora, kao što su Excel radne knjige, tekstualne/CSV datoteke, lokalne i putem interneta dostupne SQL baze podataka i/ili izvora na webu (društvenih mreža, imenika i drugih izvora na koje se možemo povezati i postavljati upite) , transformaciju jer se tako dobiveni podaci mogu očistiti, pripremiti i prilagoditi potrebama analize podataka i učitavanje („load“) u podatkovni model ili radni list.

PQ ima nekoliko korisnih značajki kao što je dodavanje novih podataka i stvaranje odnosa između različitih skupova podataka. Također možemo grupirati i sažeti podatke, a sam alat pruža mogućnost postavljanje zadanog tipa podatka u kolonama, kao i kontrolu podataka prije samog uvoza u Excel. Korištenjem PQ kao alata za povezivanje, omogućeno je dinamičko/interaktivno mijenjanje podataka čak i jednom kada podatke transformiramo u Excel tablice.



Slika 1.ETL proces u Power Query

Četiri su faze korištenja Power Query alata:

- **Povezivanje** - povezivanje s podacima u oblaku, na servisu ili lokalno
- **Transformiranje** - oblikovati podatke da odgovaraju potrebama analize
- **Kombiniranje -** integracija podataka iz više izvora radi jedinstvenog prikaza podataka
- **Učitavanje** - Učitavanje upita na radni list ili podatkovni model s mogućnošću povratka u PQ u svakom trenutku radi osvježavanja modela

Povezivanje podataka

Jedna od najvećih prednosti korištenja PQ je mogućnost povezivanja / uvoza podataka iz različitih izvora i različitih formata (xlsx, txt, folder, sharepoint, base podataka, web, email…)

Transformiranje

Omogućava oblikovanje podataka za potrebe analize. U ovoj fazi korisniku su na raspolaganju različite funkcionalnosti kao što se : pregled i postavljanje željenih tipova podataka u stupcima, dodavanje novih stupaca i formula, spajanje ili razdvajanje stupaca…

Ponekad pri uvoz tekstualnih datoteka, Excel datum pročita kao broj dana od 01.01.1900, a Power Query omogućava da se da pogreška ispravi i postavi odgovarajući tip datuma prije učitavanja podataka u Excel tablice odnosno radne listove.

Još jedna prednost PQ alata je dodavanje, odnosno kreiranje novih stupaca koji mogu nositi određene tipove podataka ili formula.

Kombiniranje

Tablica 1.

Engleski	**Hrvatski**
Sales	Prodaja
Row_Id	Jedinstveni identifikator retka
Order_Id	Jedinstveni identifikator narudžbe
Order_date	Datum narudžbe
Ship_date	Datum otpreme
Ship_mode	Vrsta otpreme
Customer_ID	Jedinstveni identifikator kupca
Cust_first_name	Ime kupca
Cust_last_name	Prezime kupca
Cust_full_name	Puno ime i prezime kupca
Segment	Segment kupca (specifičan fragment kojem kupac pripada)
Country	Država
City	Grad
State	Država (u ovo slučaju kao dio SAD-a)
Postal_code	Poštanski broj
Region	Regija
Product_ID	Jedinstvena oznaka proizvoda
Product_name	Naziv proizvoda
Category	Kategorija proizvoda
Sub_category	Pod-kategorija proizvoda
Sales	Ukupna prodaja
Quantity	Količina
Discount	Popust u %
Unit_price	Cijena po komadu
Discount $	Iznos popusta
Profit	Zarada

Pojmovnik tablice Sales – izvor autor

PQ kombinira više različitih izvora podataka tako da ih spaja ili dodaje. Operacije dodavanje i spajanje izvode se na bilo kojem upitu s tabličnim oblikom i neovisne su o izvorima podataka iz kojeg podaci dolaze.
Operacija spajanja stvara novi upit iz dva postojeća upita. Ovaj upit sadrži sve stupce iz primarne tablice, a jedan stupac služi kao navigacijska veza na povezanu tablicu. Povezana tablica sadrži sve retke koji se podudaraju sa svakim retkom iz zajedničke vrijednosti stupca u primarnoj tablici.

Učitavanje podataka

PQ omogućuje učitavanje podataka u različite Excel formate

- Excel tablicu

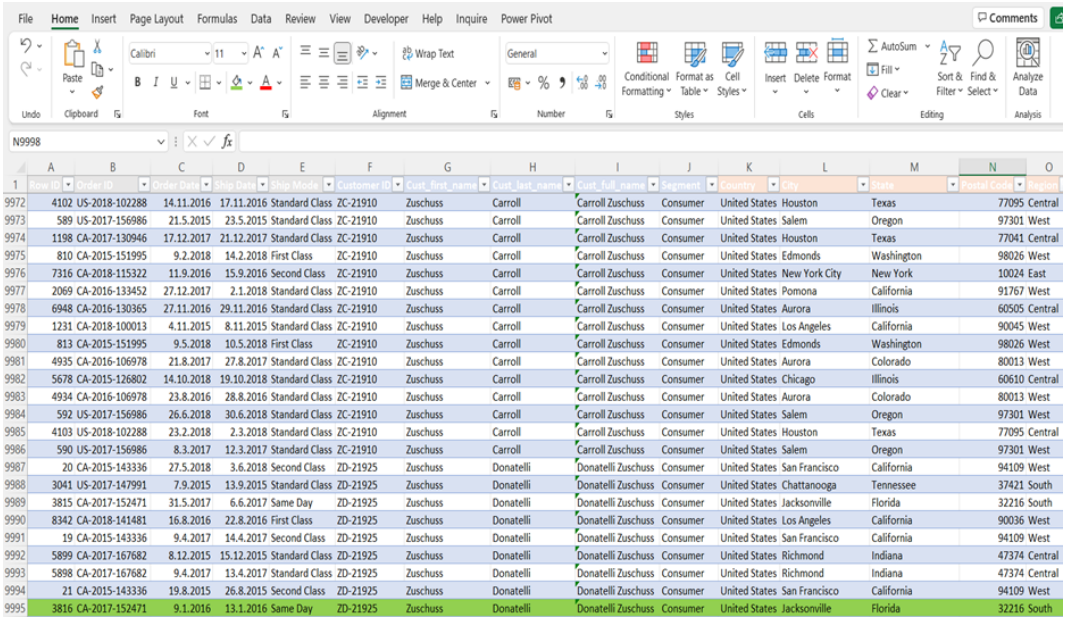
- Pivot tablicu

- Pivot grafikon

- ili stvaranje veze/konekcije na podatke (u slučaju povezivanja podataka iz više izvora)

4.1.1.1 Praktični primjer korištenja Power Query mehanizma

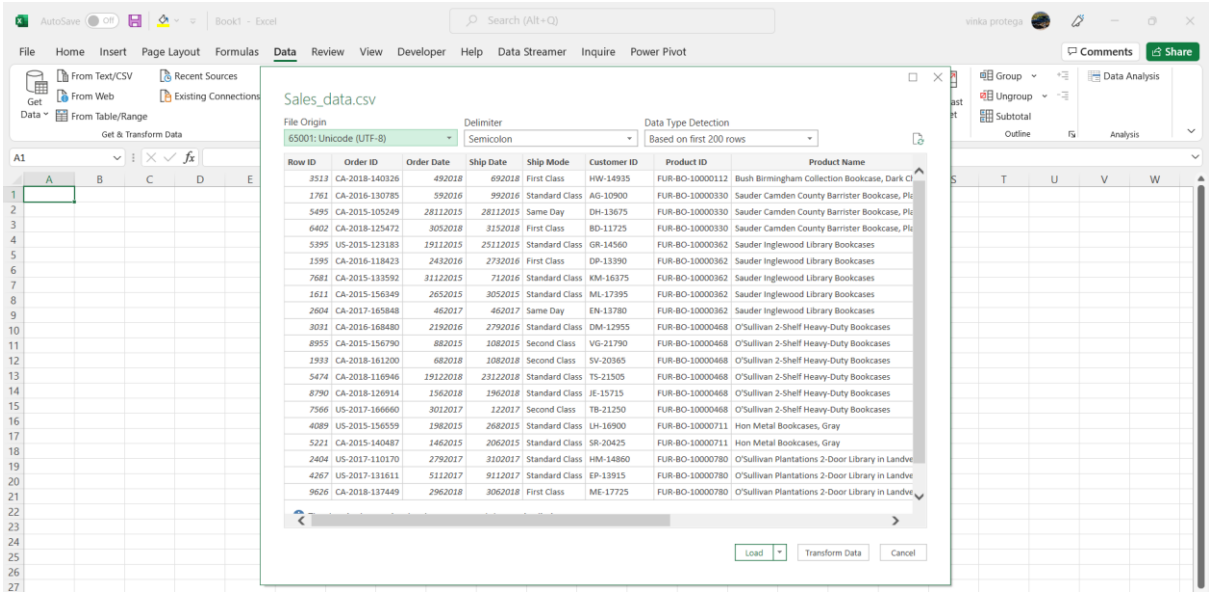
Datoteka koja se koristi u praktičnim primjerima zove se Sales.xlsx i preuzeta je sa internetske stranice www.simplilearn.com. Pojmovi u tablici su na engleskom jeziku i kao takvi bit će temelj za analizu podataka. Radi lakšeg snalaženja u pojmovima tablice Sales.xlsx, prije početka praktične primjene, u nastavku slijedi popis pojmova na engleskom i hrvatskom jeziku.
Tablica 1.



Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Cust_first_name	Cust_last_name	Cust_full_name	Segment	Country	City	State	Postal Code	Region
4102	US-2018-102288	14.11.2016	17.11.2016	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Houston	Texas	77095	Central
589	US-2017-156986	21.5.2015	23.5.2015	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Salem	Oregon	97301	West
1198	CA-2017-130946	17.12.2017	21.12.2017	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Houston	Texas	77041	Central
810	CA-2015-151995	9.2.2018	14.2.2018	First Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Edmonds	Washington	98026	West
7316	CA-2018-115322	11.9.2016	15.9.2016	Second Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	New York City	New York	10024	East
2069	CA-2016-133452	27.12.2017	2.1.2018	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Pomona	California	91767	West
6948	CA-2016-130365	27.11.2016	29.11.2016	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Aurora	Illinois	60505	Central
1231	CA-2018-100013	4.11.2015	8.11.2015	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Los Angeles	California	90045	West
813	CA-2015-151995	9.5.2018	10.5.2018	First Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Edmonds	Washington	98026	West
4935	CA-2016-106978	21.8.2017	27.8.2017	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Aurora	Colorado	80013	West
5678	CA-2015-126802	14.10.2018	19.10.2018	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Chicago	Illinois	60610	Central
4934	CA-2016-106978	23.8.2016	28.8.2016	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Aurora	Colorado	80013	West
592	US-2017-156986	26.6.2018	30.6.2018	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Salem	Oregon	97301	West
4103	US-2018-102288	23.2.2018	2.3.2018	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Houston	Texas	77095	Central
590	US-2017-156986	8.3.2017	12.3.2017	Standard Class	ZC-21910	Zuschuss	Carroll	Carroll Zuschuss	Consumer	United States	Salem	Oregon	97301	West
20	CA-2015-143336	27.5.2018	3.6.2018	Second Class	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	San Francisco	California	94109	West
3041	US-2017-147991	7.9.2015	13.9.2015	Standard Class	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	Chattanooga	Tennessee	37421	South
3815	CA-2017-152471	31.5.2017	6.6.2017	Same Day	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	Jacksonville	Florida	32216	South
8342	CA-2018-141481	16.8.2016	22.8.2016	First Class	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	Los Angeles	California	90036	West
19	CA-2015-143336	9.4.2017	14.4.2017	Second Class	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	San Francisco	California	94109	West
5899	CA-2017-167682	8.12.2015	15.12.2015	Standard Class	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	Richmond	Indiana	47374	Central
5898	CA-2017-167682	9.4.2017	13.4.2017	Standard Class	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	Richmond	Indiana	47374	Central
21	CA-2015-143336	19.8.2015	26.8.2015	Second Class	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	San Francisco	California	94109	West
3816	CA-2017-152471	9.1.2016	13.1.2016	Same Day	ZD-21925	Zuschuss	Donatelli	Donatelli Zuschuss	Consumer	United States	Jacksonville	Florida	32216	South

Slika 2.Excel tablica s 9995 redaka – izvor www.simplelearn.com

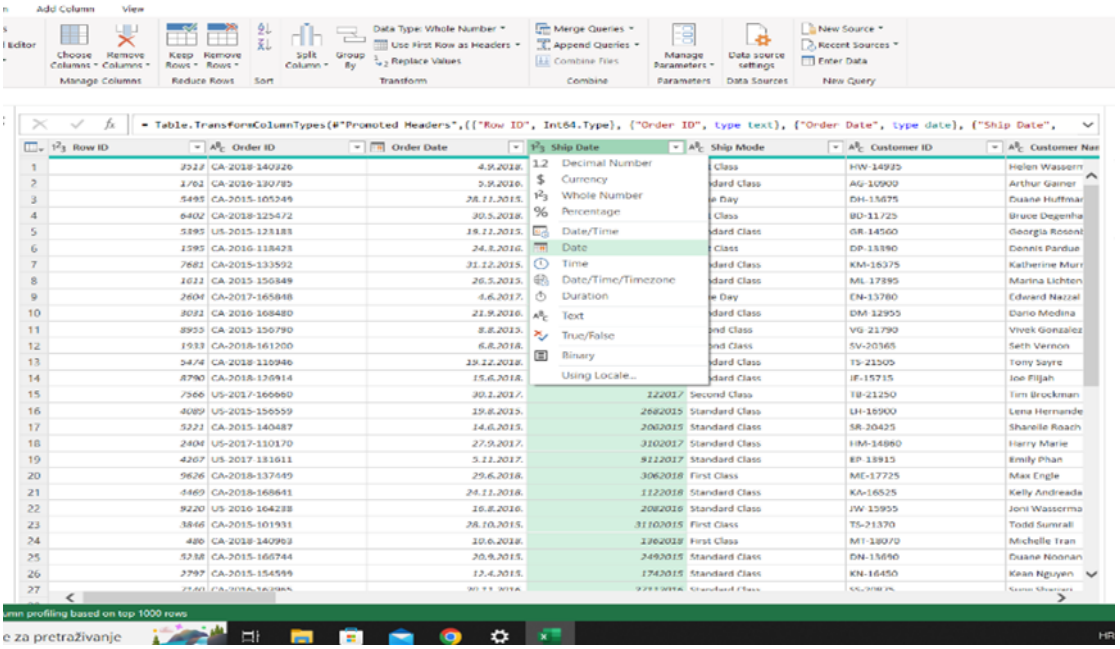
Radi bolje vizualizacije i realnijeg prikaza funkcionalnosti u PQ, izvorna tablica podijeljena je na dvije tablice, tablicu Sales_data.csv i tablicu Customer_Masters.csv. U prvom koraku prikazano je učitavanje tablice Sales_data u PQ .



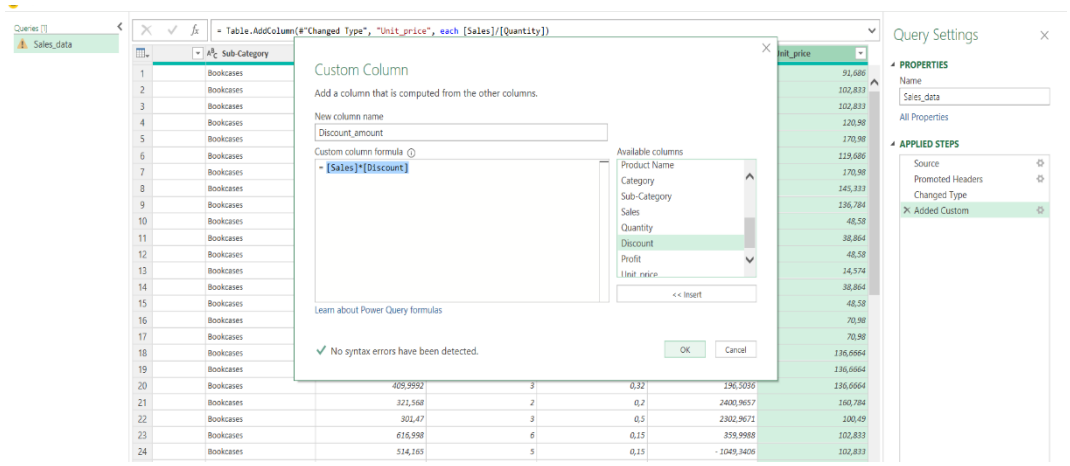
Slika 3. Učitavanje tablice sales_data.csv u PQ.

Jednom kad je tablica učitana u PQ moguće su različite manipulacije s podacima. U prikazanoj tablici, u kolonama Order_date i Ship_date datum se nalazi u neodgovarajućem formatu (prikazuje broj dana od 01.01.1900). PQ omogućava jednostavnu izmjenu i odabir željenog

formata datuma, kao i izmjenu vrste podataka u ostalim kolonama.

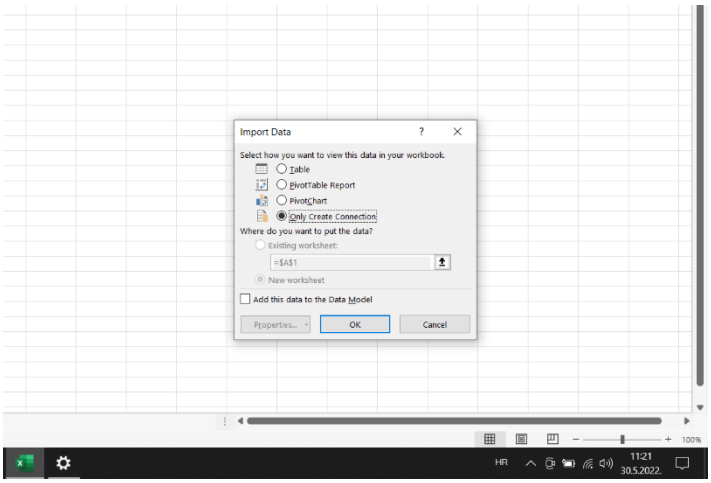


Slika 4. Postavljanje željenog formata datuma – izvor autor



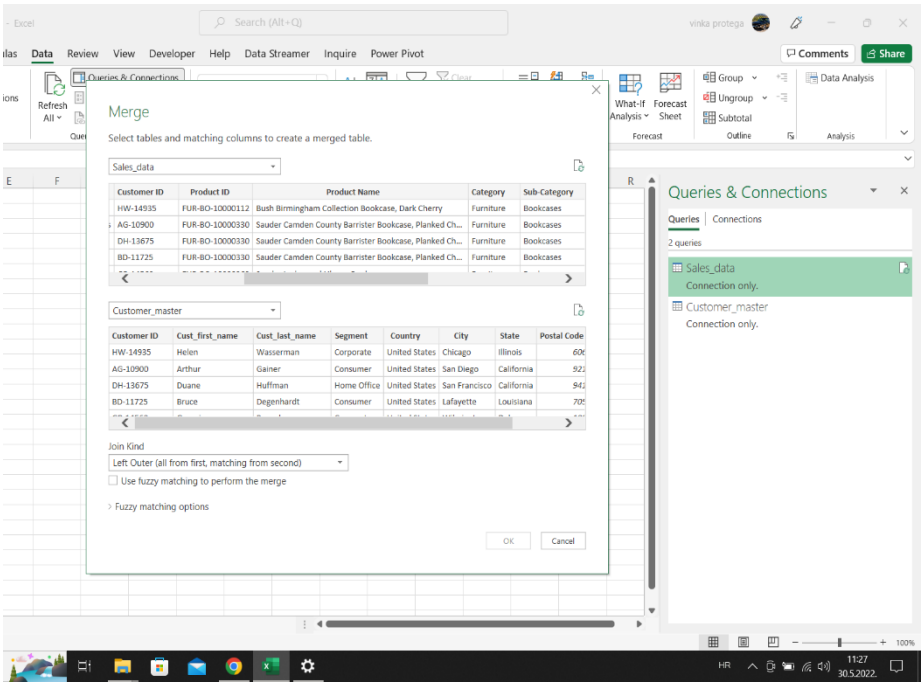
Slika 5. Dodavanje nove kolone uz korištenje formule za izračun iznosa

Za povezivanje podataka iz više izvora u PQ kreiramo veze (konekcije) na tablicu.



Slika 6. Stvaranje veza na tablicu

Datoteka Customer_masters učitana je na isti način i za tu tablicu kreirana je konekcija. Odabirom opcije MERGE, omogućeno je spajanje tablica po zajedničkom ključu (jednom ili složenom) .



Slika 7. Povezivanje tablica putem MERGE opcije

Jednom spojene tablice spremne su za daljinu analizu.

4.1.2 Analiza podataka korištenjem built in funkcija

Excel ima stotine funkcija i usklađivanja odgovarajuće formule s pravom vrstom analize podataka može biti vrlo zahtjevno. Nije nužno da najčešće i najkorisnije funkcije budu teške.

Najčešće funkcije za analizu podataka u Excelu:

CONCAT ()

Sintaksa: = CONCAT(text1,text2)

Prilikom provođenja analize podataka, formula =CONCAT jedna je od najjednostavnijih za razumijevanje, ali najmoćnija. Tekst, brojevi, datumi i drugi podaci iz brojnih ćelija mogu se kombinirati u jednu ćeliju.

SUM ()

Sintaksa = SUM (num1;num2..)

Funkcija koja zbraja pojedinačne vrijednosti, reference ćelija ili raspone ili kombinaciju svo troje.

SUMPRODUCT ()

Sintaksa = SUMPRODUCT(RANGE1,RANGE2)

Funkcija množi jedan raspon vrijednosti s odgovarajućim recima.

COUNTA()

Sintaksa = COUNTA(value1, [value2], ...)

Funkcija broji ćelije koje nisu prazne u zadanom rasponu.

RIGHT/LEFT/MID ()

Sintaksa: = LEFT /RIGHT(text, [num_chars])

= MID (text, **start_position**, number_of_characters)

Sintaksa za funkcije 'LEFT' i 'RIGHT' potpuno je ista dok funkcija 'MID' sadrži jedan dodatni parametar.

Funkcije 'LEFT', 'MID' i 'RIGHT' osmišljene su kako bi nam omogućile dohvaćanja pod-niza iz jednog niza podataka na temelju relativnog položaja tog podniza unutar izvornog niza.

YEAR/DAY/MONTH (

Sintaksa = YEAR/DAY/MONTH(from_date)
Funkcija izdvaja dan, mjesec ili godinu iz datuma.

MAX ()

Sintaksa = MAX(number1, [number2], …)
Funkcija MAX spada u kategoriju statističkih funkcija u Excelu. MAX će vratiti najveću vrijednost na danom **popisu argumenata.** Iz zadanog skupa brojčanih vrijednosti vratit će najvišu vrijednost.
U financijskoj analizi, MAX može biti koristan u izračunu najviše ocjene, najbržeg vremena, najvećeg iznosa troškova ili prihoda itd.

MIN ()

Sintaksa = MIN (number1, [number2], ...)
Funkcija MIN vraća najmanju brojčanu vrijednost u navedenim podacima. Funkcija MIN može se koristiti za vraćanje najmanje vrijednosti iz bilo koje vrste brojčanih podataka. Na primjer, MIN može vratiti najbrže vrijeme u utrci, najraniji datum, najmanji postotak, najnižu temperaturu ili najniži broj prodaje.
Funkcija MIN uzima više argumenata u obliku broj1, broj2, broj 3, itd. do ukupno 255.

VLOOKUP ()

Sintaksa = VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])
Jedna od najkorisnijih i najčešće korištenih funkcija kod analize podataka u Excelu. VLOOKUP funkcija skenira tablicu, pronalazi podudaranje i vraća točan rezultat.
VLOOKUP je funkcija programa Excel za dobivanje podataka iz tablice organizirane **okomito, odnosno vertikalno (VLOOKUP-vertical lookup)**
VLOOKUP uzima četiri argumenta: lookup_value, table_array, column_index_num i range_lookup.
Lookup_value je vrijednost koju treba tražiti, a table_array raspon je okomitih podataka koje

treba pogledati.

Prvi stupac **table_array** mora sadržavati tražene vrijednosti za pretraživanje.

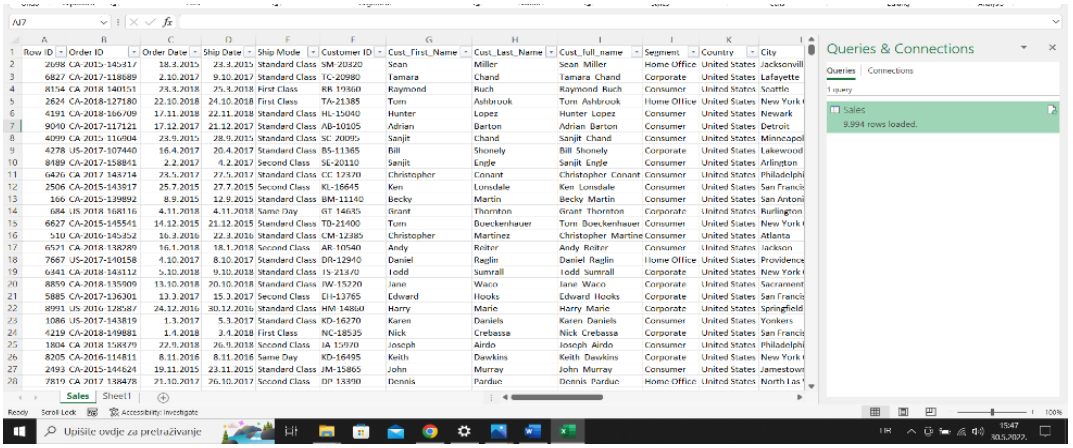
Argument **column_index_num** broj je stupca vrijednosti za dohvaćanje, gdje je prvi stupac table_array stupac 1.

Konačno, **range**_lookup kontrolira ponašanje podudaranja. Ako je range_lookup TRUE, odnosno 1, VLOOKUP će izvesti približno podudaranje. Ako je range_lookup FALSE, odnosno 0, VLOOKUP će izvesti točno podudaranje.

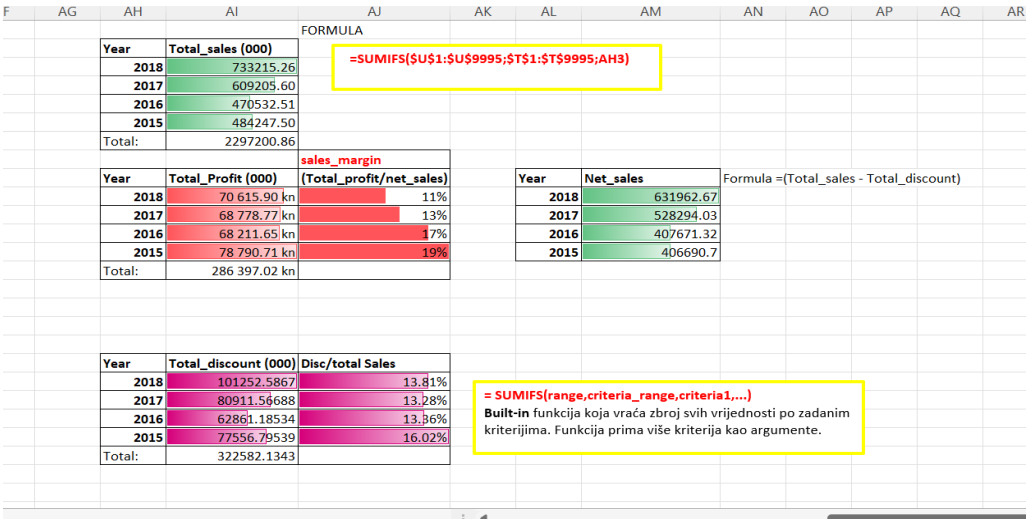
4.1.2.1 Praktična primjer primjene analize podataka primjenom Excel funkcija

Kao je već navedeno, Excel sadrži na stotine funkcija koje korisniku omogućuju različite dohvate, izračune i trenutne uvide. U praktičnom primjeru kreirani su poslovni izvještaji korištenjem Excel built-in funkcija i uvjetnog oblikovanja (conditional formatting).

U ovo primjeru provedena je analiza podataka maloprodaje tvrtke xy. Izvorne datoteke bile su u csv. formatu. Uvoz u Excel napravljen je korištenjem PQ alata. Podaci se nalazili u dva odvojena dokumenta Sales_data i Customer_masters. Kod transformacije podataka postavljen je ispravan tip podatka za datum (ispravljena greška pri uvozu u Excel), radi zahtjeva analize i izrade izvještaja, dodani su novi stupci s formulama (Unit Price, Discount_amount) a radi planirane obrade podataka za SQL, stupac Customer razdvojen je na Cast First_name i Cust Last name. Obje tablice spojene su putem istoznačnog stupca Customer_ID i uvezene u Excel.



Slika 8. Tablica sales



FORMULA

=SUMIFS($U$1:$U$9995;$T$1:$T$9995;AH3)

Year	Total_sales (000)
2018	733215.26
2017	609205.60
2016	470532.51
2015	484247.50
Total:	2297200.86

sales_margin

Year	Total_Profit (000)	(Total_profit/net_sales)
2018	70 615.90 kn	11%
2017	68 778.77 kn	13%
2016	68 211.65 kn	17%
2015	78 790.71 kn	19%
Total:	286 397.02 kn	

Year	Net_sales
2018	631962.67
2017	528294.03
2016	407671.32
2015	406690.7

Formula =(Total_sales - Total_discount)

Year	Total_discount (000)	Disc/total Sales
2018	101252.5867	13.81%
2017	80911.56688	13.28%
2016	62861.18534	13.36%
2015	77556.79539	16.02%
Total:	322582.1343	

= SUMIFS(range,criteria_range,criteria1,...)

Built-in funkcija koja vraća zbroj svih vrijednosti po zadanim kriterijima. Funkcija prima više kriterija kao argumente.

Slika 9. Izvještaj prodaje - Sales Report -izvor autor

Najveću ukupnu prodaju tvrtka je ostvarila 2018. godine u iznosu od 733.215.260,00 $, ali ipak najveći ukupni profit ostvaren je 2015. u iznosu 78,790.709,00$.

Ovo je odličan primjer izvještaja koji osim što odgovara na postavljena pitanja, otvara i prostor za detaljniji uvid u poslovanje.

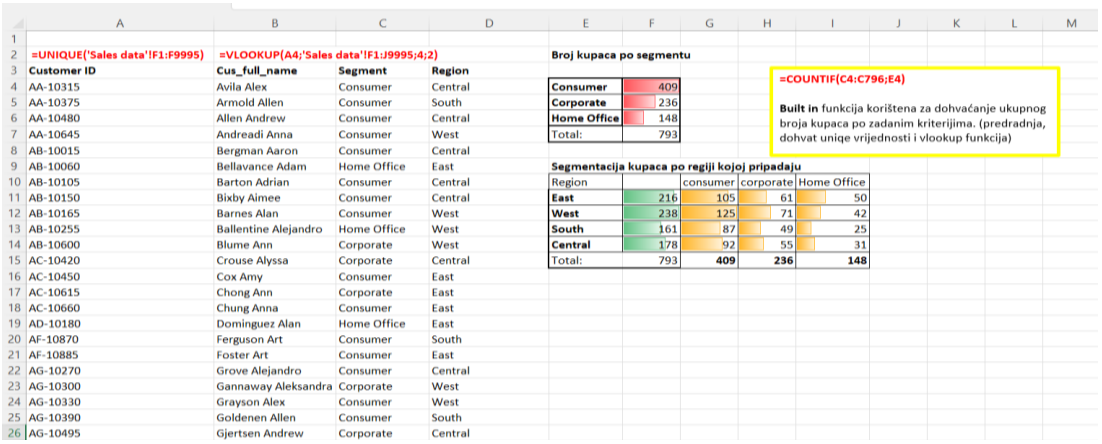
Analiza iznosa popusta po godinama otklanjanju mogućnost da je to razlog za smanjenje profita, jer iako je najveći iznos popusta ostvaren 2018. godine, njegov udio u ukupnoj prodaji relativno je jednak kroz sva promatrana razdoblja i čak nešto veći u 2015. godini, kada je tvrtka ostvarila najveći profit.

Trgovačka marža(Sales_Margin) izvedena je kao omjer profita i ukupne neto prodaje (ukupna prodaja – ukupan popust). Iz izvještaja je vidljivo da se trgovačka marža s godinama smanjuje te je 2018. iznosila 11% što je za 8% manje nego 2015 godine kada je trgovačka marža bila visokih 19%.

- **Analiza kupaca**

Analiza i segmentacija kupaca čine važan dio poslovnog izvještavanja. U ovom primjeru

korištene su tri jednostavne Excel built-in funkcije; UNIQUE za dohvat jedinstvenih vrijednosti, VLOOKUP za dohvat imena kupaca, segmenta i regija te COUNTIF funkcija za dohvat ukupnog broja kupaca za pojedini segment i regiju.



=UNIQUE('Sales data'!F1:F9995) | =VLOOKUP(A4;'Sales data'!F1:J9995;4;2)

Customer ID	Cus_full_name	Segment	Region
AA-10315	Avila Alex	Consumer	Central
AA-10375	Armold Allen	Consumer	South
AA-10480	Allen Andrew	Consumer	Central
AA-10645	Andreadi Anna	Consumer	West
AB-10015	Bergman Aaron	Consumer	Central
AB-10060	Bellavance Adam	Home Office	East
AB-10105	Barton Adrian	Consumer	Central
AB-10150	Bixby Aimee	Consumer	Central
AB-10165	Barnes Alan	Consumer	West
AB-10255	Ballentine Alejandro	Home Office	West
AB-10600	Blume Ann	Corporate	West
AC-10420	Crouse Alyssa	Corporate	Central
AC-10450	Cox Amy	Consumer	East
AC-10615	Chong Ann	Corporate	East
AC-10660	Chung Anna	Consumer	East
AD-10180	Dominguez Alan	Home Office	East
AF-10870	Ferguson Art	Consumer	South
AF-10885	Foster Art	Consumer	East
AG-10270	Grove Alejandro	Consumer	Central
AG-10300	Gannaway Aleksandra	Corporate	West
AG-10330	Grayson Alex	Consumer	West
AG-10390	Goldenen Allen	Consumer	South
AG-10495	Gjertsen Andrew	Corporate	Central

Broj kupaca po segmentu

Consumer	409
Corporate	236
Home Office	148
Total:	793

Segmentacija kupaca po regiji kojoj pripadaju

Region		consumer	corporate	Home Office
East	216	105	61	50
West	238	125	71	42
South	161	87	49	25
Central	178	92	55	31
Total:	793	409	236	148

=COUNTIF(C4:C796;E4)

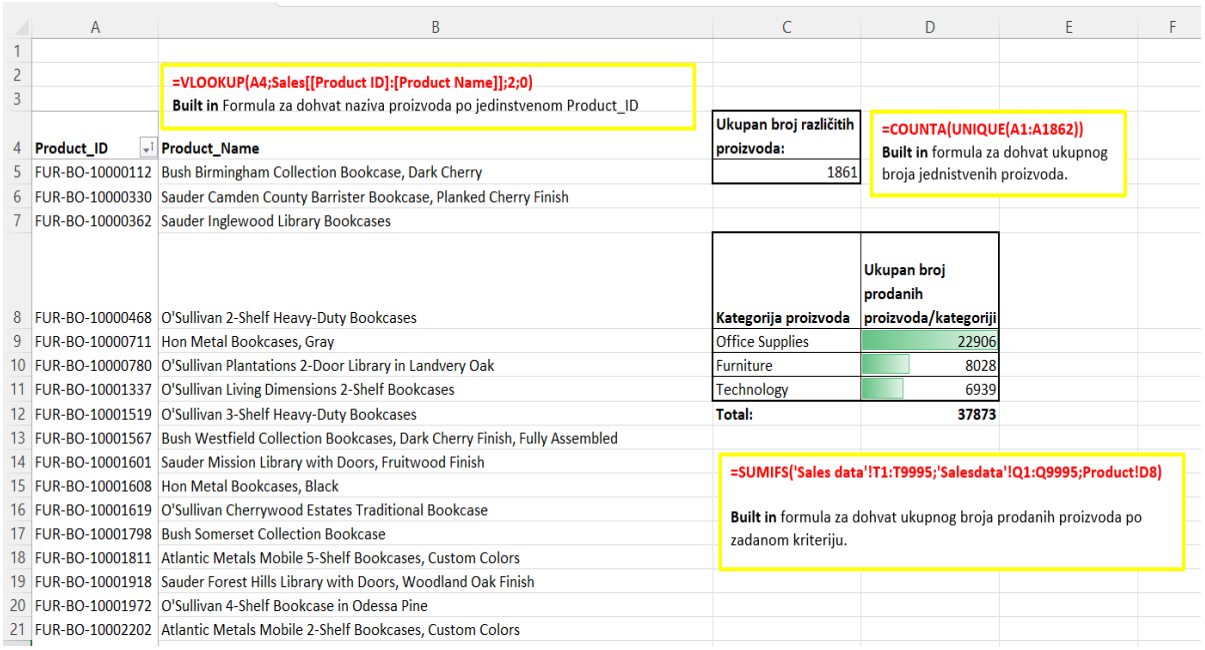
Built in funkcija korištena za dohvaćanje ukupnog broja kupaca po zadanim kriterijima. (predradnja, dohvat uniqe vrijednosti i vlookup funkcija)

Slika 10. Izvještaj analiza kupaca- Customer report -izvor autor

Analizom podataka o kupcima tvrtka je dobila informacije o ukupnom broju kupaca, odnosno 793 jedinstvena kupca. Najveći broj kupaca pripada segmentu Consumer, a najviše kupaca iz tog segmenta se nalazi u zapadnoj regiji. Najmanji broj kupaca pripada Home Office segmentu.

- **Analiza proizvoda**

Izvještaj o analizi proizvoda govori nam koliko različitih proizvoda kompanija prodaje, u koliko kategorija i koliko je ukupno prodano proizvoda u određenoj kategoriji.



=VLOOKUP(A4;Sales[[Product ID]:[Product Name]];2;0)

Built in Formula za dohvat naziva proizvoda po jedinstvenom Product_ID

Product_ID	Product_Name
FUR-BO-10000112	Bush Birmingham Collection Bookcase, Dark Cherry
FUR-BO-10000330	Sauder Camden County Barrister Bookcase, Planked Cherry Finish
FUR-BO-10000362	Sauder Inglewood Library Bookcases
FUR-BO-10000468	O'Sullivan 2-Shelf Heavy-Duty Bookcases
FUR-BO-10000711	Hon Metal Bookcases, Gray
FUR-BO-10000780	O'Sullivan Plantations 2-Door Library in Landvery Oak
FUR-BO-10001337	O'Sullivan Living Dimensions 2-Shelf Bookcases
FUR-BO-10001519	O'Sullivan 3-Shelf Heavy-Duty Bookcases
FUR-BO-10001567	Bush Westfield Collection Bookcases, Dark Cherry Finish, Fully Assembled
FUR-BO-10001601	Sauder Mission Library with Doors, Fruitwood Finish
FUR-BO-10001608	Hon Metal Bookcases, Black
FUR-BO-10001619	O'Sullivan Cherrywood Estates Traditional Bookcase
FUR-BO-10001798	Bush Somerset Collection Bookcase
FUR-BO-10001811	Atlantic Metals Mobile 5-Shelf Bookcases, Custom Colors
FUR-BO-10001918	Sauder Forest Hills Library with Doors, Woodland Oak Finish
FUR-BO-10001972	O'Sullivan 4-Shelf Bookcase in Odessa Pine
FUR-BO-10002202	Atlantic Metals Mobile 2-Shelf Bookcases, Custom Colors

Ukupan broj različitih proizvoda:
1861

=COUNTA(UNIQUE(A1:A1862))

Built in formula za dohvat ukupnog broja jednistvenih proizvoda.

Kategorija proizvoda	Ukupan broj prodanih proizvoda/kategoriji
Office Supplies	22906
Furniture	8028
Technology	6939
Total:	**37873**

=SUMIFS('Sales data'!T1:T9995;'Salesdata'!Q1:Q9995;Product!D8)

Built in formula za dohvat ukupnog broja prodanih proizvoda po zadanom kriteriju.

Slika 11.Izvještaj analize proizvoda – Products report -izvor autor

Iz izvještaja je vidljivo da kompanija prodaje ukupno 1861 različiti proizvod, a najveću prodaju ostvaruje u kategoriji Office Supplies.

4.1.3 Analiza podataka korištenjem zaokretnih tablica i grafikona

Zaokretne ili pivot tablice
Daleko najčešći alat za analizu podataka je zaokretna ili pivot tablica. To je napredan alat koji sažima tisuće redova podataka u koristan i pregledan izvještaj . Omogućuje korisnicima različite manipulacije s podacima kao što su uporaba različitih funkcija sažimanja (SUM), prikaz vremenskih dimenzija podataka, pregled i usporedbu podataka i vizualizaciju, odnosno izradu različitih pivot grafikona.
Pivot tablica služi za detaljniji uvid u podatke . Podatke u zaokretnu tablicu možemo učitati putem Power Query alata ili odabirom tipke **Umetanje (Insert)** na alatnoj traci.
Analizu podataka, pivot tablice omogućuju izvođenjem triju operacija:

a) **grupiranje podataka u kategorije – tablica automatski grupira velike količine** podataka u manje dijelove kojima se lakše upravlja.

b) **sažimanje podataka** - u kombinaciji s grupiranjem podataka prema jedinstvenim vrijednostima u jednom ili više polja, pivot tablica također prikazuje sažete izračune za svaku grupu. Zadani izračun je zbroj, što znači da za svaku grupu Excel zbraja sve vrijednosti u određenom polju. Na primjer, ako podaci imaju polje Regija i polje Prodaja, zaokretna tablica može grupirati jedinstvene vrijednosti regije i prikazati ukupne vrijednosti prodaje za svaku od regija. Moguće je koristiti i druge zbirne izračune uključujući brojanje, prosjek, maksimum, minimum i standardnu devijaciju. Ako pretpostavimo da podaci o prodaji također imaju polje proizvoda. Moguće je postaviti zaokretnu tablicu za prikaz ukupne prodaje za svaki proizvod, raščlanjene po regijama.

c) **filtriranje podataka** kako bi se prikazali samo zapisi s kojima želimo raditi pivot tablica omogućuje pregled samo podskupa podataka. Na primjer, prema zadanim postavkama, grupiranje zaokretne tablice prikazuje sve jedinstvene vrijednosti u polju. Međutim, moguće je manipulirati svakim grupiranjem kako bi sakrili vrijednosti koje ne želimo vidjeti, a prikazali samo one vrijednosti koje su nam važne za kreiranje izvješća. Svaka zaokretna tablica također dolazi s filtrom izvješća koji omogućuje primjenu filtera na cijelu zaokretnu tablicu. Na primjer, pretpostavimo da podaci o prodaji također uključuju polje segment kupca. Postavljanjem ovog polja u filtar, moguće je filtrirati izvješće zaokretne tablice da prikazuje rezultate za samo jedan segment kupaca.

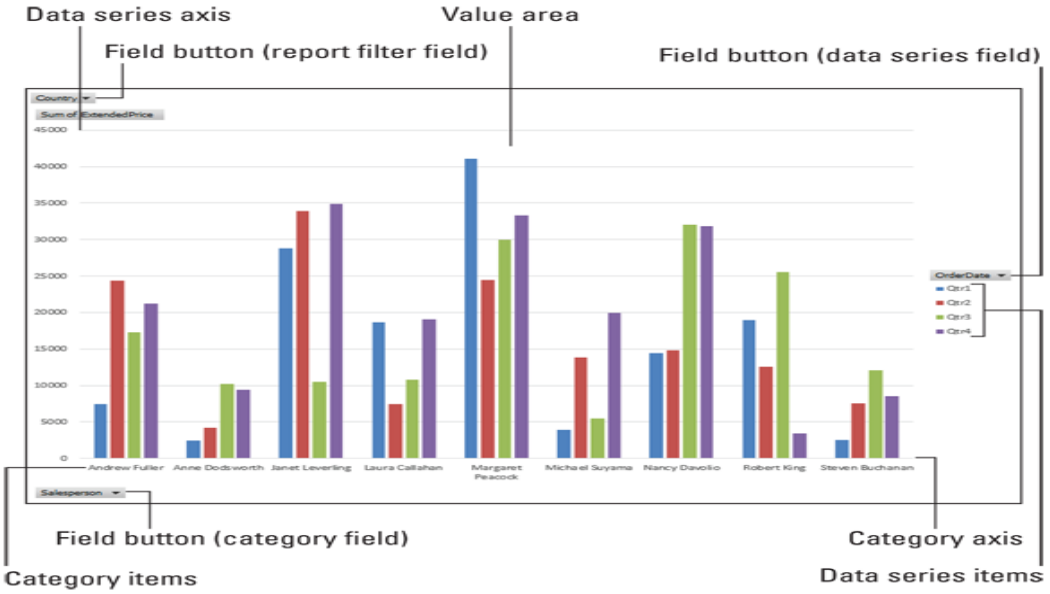
Pivot grafikoni
Pivot grafikoni **naziva se i** zaokretni grafikon**. D**olazi s većinom istih značajki koje zaokretne tablice čine tako moćnim: možete filtrirati rezultate da biste vidjeli samo podatke koji su vam potrebni**. Pivot grafikon sadrži sljedeća polja:**

Kategorije grafa (Chart categories) – OS X – pivot grafikoni automatski grupiraju podatke

u manje grupe. Ovo polje odgovara redcima u Pivot tablici

Polje nizova/serija podataka (Chart data series) – odgovara stupcima u pivot tablici

Polje numeričke vrijednosti (Values) – OS Y – kao i pivot tablice, ovo polje sadrži vrijednosti , odabirom postavki vrijednosti mogu biti u različitim oblicima (SUM,COUNT…)

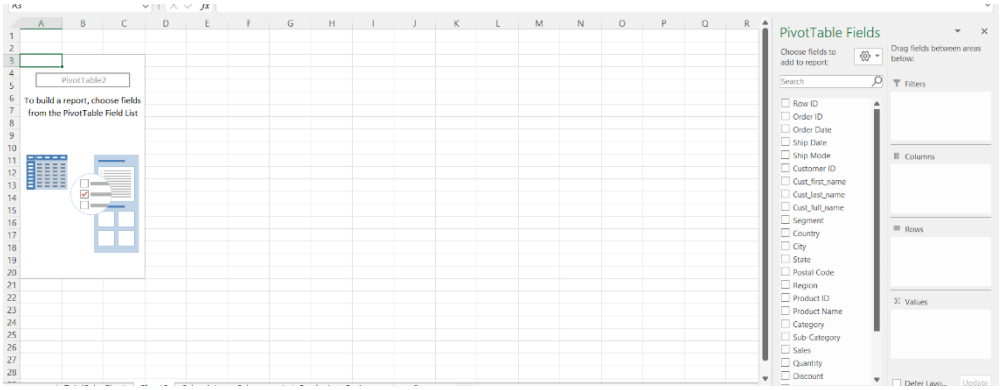


Slika 12. Polja pivot grafa - izvor "Excel data analysis for Dummies"

Pivot grafikoni dolaze sa svim prednostima pivot tablice i Excel grafikona. Kreiranje pivot grafikona iz pivot tablice vrlo je jednostavan postupak. Ipak ove vrste grafikona imaju svoje nedostatke. Postoje ograničenja u izboru tipa grafikona, pa tako tipovi poput raspršenog prikaza (scattered) i mjehurićastog (bubble) nisu mogući odabir za pivot grafikon. Također u situacijama kada je dataset velik i ima velik broj kategorija, postoji mogućnost da grafikon postane nepregledan i nečitljiv.

4.1.3.1 Praktični primjer primjene analize podataka primjenom Pivot tablica i grafikona

Pivot tablica kreirana je iz radnog lista odabirom tip opcije **Umetanje- pivot tablica.** Radi preglednijeg i jednostavnijeg rada, odabrana je opcija kreiranja pivot tablice na novom radnom listu.

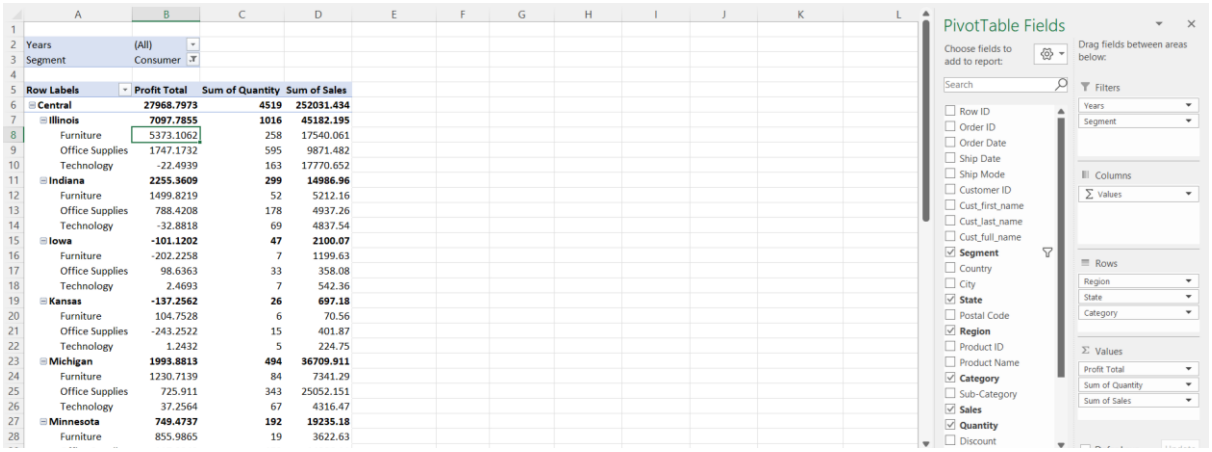


Slika 13 Kreiranje pivot tablice

U praktičnom dijelu prikazane su tri opcije/načina korištenja pivot tablica.

- Izrađena pivot tablica s filterima
- Dvije pivot tablice s promijenjenim prikazom vrijednosti
- Pivot tablica i grafikon uz korištenje opcije Slicer

Za prvi primjer kreirana je pivot tablica koja prikazuje ukupan, profit, prodaju i prodanu količinu proizvoda grupiranu po regiji, državi i kategoriji proizvoda. Kao filter tablica koristi godine i segment kupca radi detaljnijeg uvida u prodaju i profit po godinama i kroz segmentaciju kupaca.



Slika 14. Pivot tablica s odabranim poljima i filterima

Pivot tablice pružaju i mogućnost da vrijednosti u polju Values budu prikazane u obliku

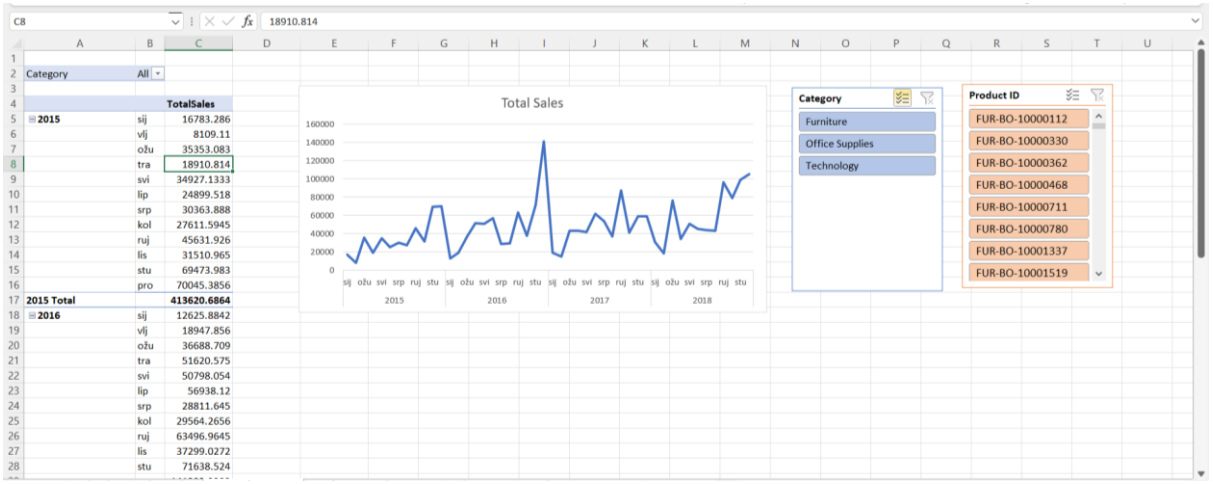
postotne promjena u odnosu na prethodnu vrijednost. U tablici lijevo prikazana je postotna promjena u ukupnoj prodaji i profitu u odnosu na prethodnu godinu, a u desnoj tablici prikazana je postotna promjena broja narudžbi odnosno broja transakcija u odnosu na prethodnu godinu.

Region	Years	Order Date	Total sales	% Change Sales	Total Profit	% Change profit
Central			501 239.89		59 278.25	
	2015		97 831.93		20 752.29	
	2016		142 547.50	45.71%	6 673.59	-67.84%
	2017		108 608.21	-23.81%	14 450.64	116.53%
	2018		152 252.25	40.18%	17 401.73	20.42%
East			678 781.24		96 582.33	
	2015		112 633.63		22 384.52	
	2016		171 076.10	51.89%	21 049.25	-5.97%
	2017		181 261.63	5.95%	28 432.66	35.08%
	2018		213 809.88	17.96%	24 715.90	-13.07%
South			391 721.91		46 877.44	
	2015		63 760.72		9 217.93	
	2016		111 529.66	74.92%	16 784.09	82.08%
	2017		85 067.43	-23.73%	12 761.56	-23.97%
	2018		131 364.10	54.42%	8 113.86	-36.42%
West			725 457.82		83 659.00	
	2015		139 394.40		17 883.92	
	2016		174 669.70	25.31%	11 971.06	-33.06%
	2017		186 013.81	6.49%	33 240.38	177.67%
	2018		225 379.91	21.16%	20 563.65	-38.14%
Grand Total			2 297 200.86		286 397.02	

Region	Years	Number of transac	Count of Order ID	% change orders
Central	2015	449	449	
	2016	505	505	12.47%
	2017	602	602	19.21%
	2018	767	767	27.41%
Central Total		2323	2323	
East	2015	576	576	
	2016	583	583	1.22%
	2017	747	747	28.13%
	2018	942	942	26.10%
East Total		2848	2848	
South	2015	325	325	
	2016	367	367	12.92%
	2017	402	402	9.54%
	2018	526	526	30.85%
South Total		1620	1620	
West	2015	643	643	
	2016	647	647	0.62%
	2017	836	836	29.21%
	2018	1077	1077	28.83%
West Total		3203	3203	
Grand Total		9994	9994	

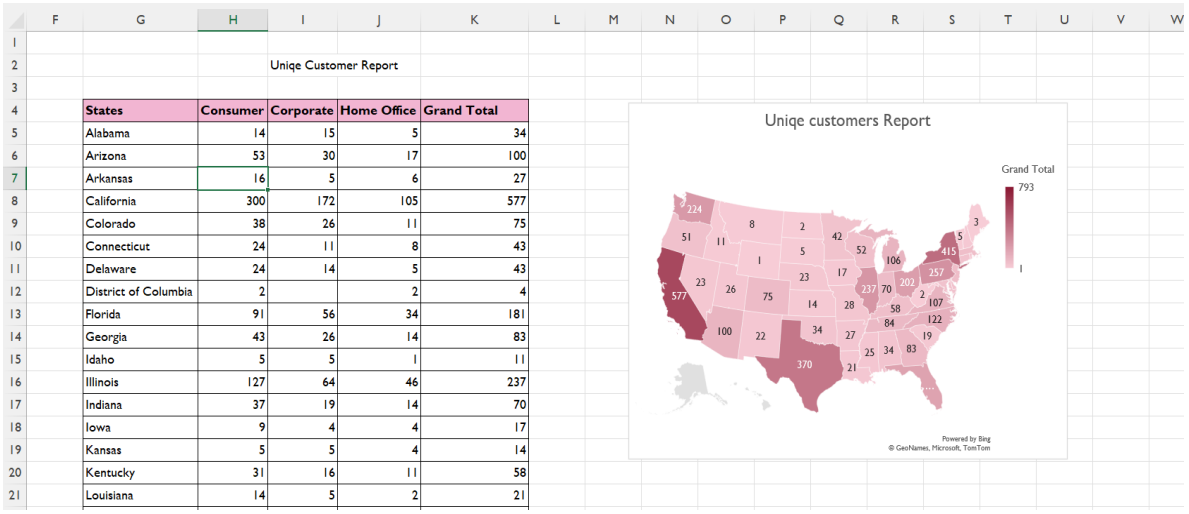
Slika 15.Usporedne pivot tablice

Pivot grafikoni imaju sva svojstva pivot tablica (dinamičnost, grupiranje i filtriranje). Korištenje ovog alata čini vizualizaciju podataka jednostavnom i preglednom. U trećem primjeru izrađen je godišnji izvještaj prodaje uz korištenje pivot tablice i vizualizaciju podatka putem pivot grafikona. Korištenje „Slicer"-a omogućava vrlo detaljan uvid u kretanje prodaje određene kategorije proizvoda, kao i dinamiku prodaje pojedinačnog proizvoda. Odabirom željenih filtera „Slicer"-a vizualizacija postaje dinamična i vrlo detaljna.



Slika 16. Izvještaj o godišnjoj prodaji proizvoda

Kako je već navedeno, pivot grafikoni i tablice imaju i svoja ograničenja. Nije moguće izraditi grafikon mapu korištenjem pivot tablice. Za izradu mape u sljedećem primjeru, podatke iz pivot tablice bilo je potrebno izvesti i tek tada oblikovati Excel grafikon koji prikazuje koliko različitih kupaca tvrtka broji u pojedinoj državi unije („Unique customer report“)



Uniqe Customer Report

States	Consumer	Corporate	Home Office	Grand Total
Alabama	14	15	5	34
Arizona	53	30	17	100
Arkansas	16	5	6	27
California	300	172	105	577
Colorado	38	26	11	75
Connecticut	24	11	8	43
Delaware	24	14	5	43
District of Columbia	2		2	4
Florida	91	56	34	181
Georgia	43	26	14	83
Idaho	5	5	1	11
Illinois	127	64	46	237
Indiana	37	19	14	70
Iowa	9	4	4	17
Kansas	5	5	4	14
Kentucky	31	16	11	58
Louisiana	14	5	2	21

Slika 17.Grafikon mapa -Unique customer report

Excel je odlična i robusna aplikacija za analizu podataka, ali ima svoj ograničenja. Jedno od značajnijih je brzina u obradi velike količine podataka. Rješenje problema Microsoft nudi u obliku SQL server management Studio.

4.2 Microsoft SQL Server Management Studio

SQL Server Management Studio (SSMS) integrirano je okruženje za upravljanje SQL infrastrukturom, od SQL Servera do Azure SQL baze podataka. SSMS sadrži alate za konfiguriranje, nadzor i administriranje instanci SQL Servera i baza podataka. Koristi se za implementaciju, nadzor i nadogradnju komponenti vezanih za podatke te za izradu upita i skripti. SSMS se najčešće upotrebljava za upite, dizajn i upravljanje bazama podataka i skladištima podataka, bilo da se isti nalaze na lokalnom računalu ili u oblaku.[3]

[3] https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver16 posjećeno 29,05,2022,

Osnovne komponente SSMS-a su:

- „Registered servers“ (registrirani poslužitelji) - omogućuju registraciju instanci „Database Engine“-a i povezivanje s njima
- „Object Explorer“ (Explorer za objekte) - sadrži prikaz svih objekata baze podataka na poslužitelju.
- Query editor – (uređivač upita) omogućuje krajnjim korisnicima generiranje, izvršavanje i pohranjivanje Transact-SQL naredbi i upita
- Solution Explorer (Explorer za rješenja) omogućuje stvaranje rješenja.

4.2.1 SQL -Structured Query Language

Programski jezik mehanizma baze podataka(„database engine“) u SSMS je Transact-SQL ili T-SQL. To je inačica odnosno proširenje standardnog SQL-a (Oracleova proširena verzija SQL-a naziva se „Procedural Language for SQL“ ili skraćeno PL/SQL).
SQL („Structured Query Language“) je standardizirani programski jezik koji se koristi za upravljanje relacijskim bazama podataka i izvođenje različitih operacija nad podacima koji se u njima nalaze. Stvoren je 1970-ih, a standardiziran je početkom 1980-tih. Službeni SQL standard usvojio je Američki nacionalni institut za standarde (ANSI) 1986, a Međunarodna organizacija za standardizaciju (ISO) usvojila je standard 1987. Nove verzije SQL standarda objavljuju se svakih nekoliko godina, posljednja 2016.
SQL se koristi ne samo za administraciju baza podataka, već i za izradu skripti za integraciju podataka i te za analizu podataka, odnosno kreiranje analitičkih upita.[4]
Pet je osnovnih skupina naredbi u SQL -u:

1. DDL - Data Definition Language – naredbe za definiranje odnosno kreiranje podatkovnih tablica
2. DML – Data Manipulation Language - koriste se za manipulaciju podacima u postojećim tablicama dodavanjem, mijenjanjem ili uklanjanjem podataka
3. TCL -Transaction Control Language - naredbe koje se koriste za promjenu stanja nekih podataka -- na primjer, COMMIT ili ROLLBACK .
4. DQL – Dana Query Language –(SELECT) -naredbe za stvaranje upita(često se

[4] https://www.techtarget.com/searchdatamanagement/definition/SQL posjećeno 03.06.2022,

navodi pod DML)

5. Data Control Language – naredbe se koriste za dodjelu ili opoziv korisničkih povlastica pristupa.(GRANT, REVOKE)

Ad 1) DDL – koristi sljedeće naredbe za definiranje podatkovnih tablica:

CREATE - naredba se koristi za stvaranje baze podataka ili njezinih objekata (kao što su tablica, indeks, funkcija, pogledi, procedura pohranjivanja i okidači).

DROP - naredba se koristi za brisanje objekata iz baze podataka.

ALTER - za promjenu strukture baze podataka.

TRUNCATE – naredba koristi se za uklanjanje svih zapisa iz tablice, uključujući i uklanjanje svih prostora dodijeljenih zapisima.

COMMENT – je naredba koja se koristi za dodavanje komentara

RENAME - koristi se za preimenovanje objekta koji postoji u bazi podataka.

Ad 2) DML – koristi naredbe za modeliranje podataka u tablicama i to:

INSERT –koristi se za umetanje podataka u tablicu.

UPDATE - za ažuriranje postojećih podataka unutar tablice.

DELETE - za brisanje zapisa iz tablice baze podataka

LOCK – za kontrolu dostupnosti podataka iz tablica

Ad 3) TCL – naredbe za kontrolu izvršenja transakcija su.

COMMIT -izvršava transakciju.

ROLLBACK - vraća transakciju unatrag u slučaju bilo kakve pogreške.

SAVEPOINT- postavlja točku spremanja unutar transakcije.

SET TRANSACTION - navedite karakteristike transakcije.

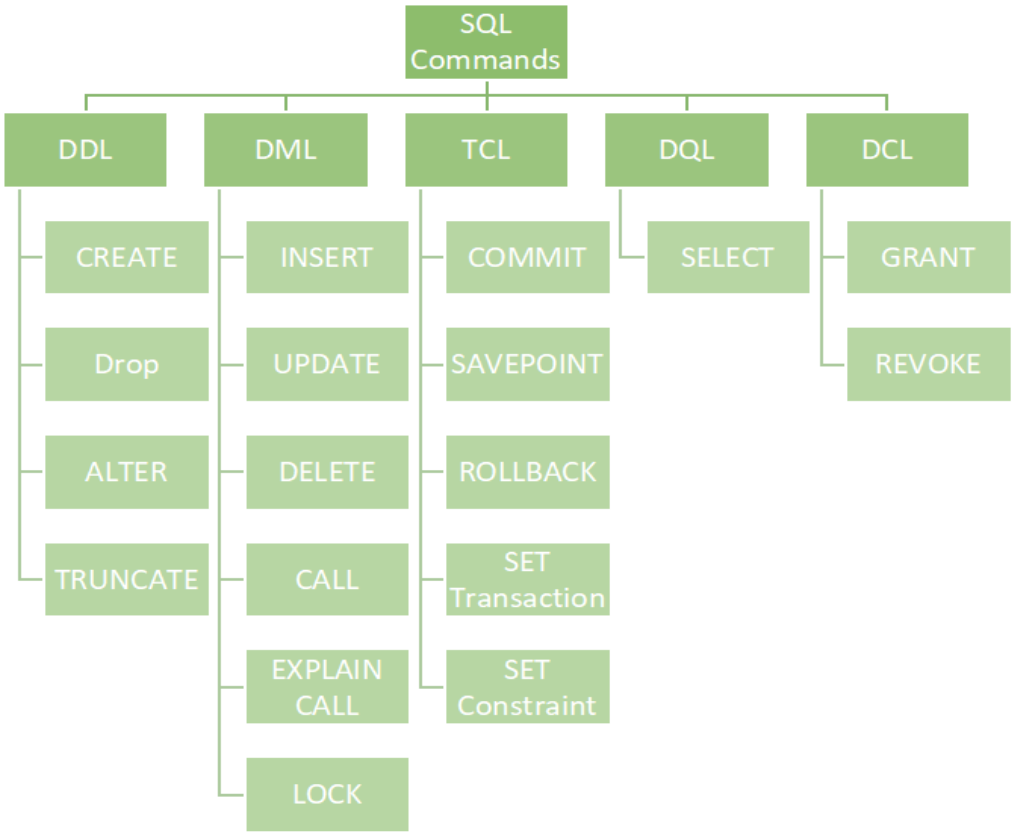
Ad 4) DQL koristi samo jednu naredbu

SELECT – kojom se kreira svaki upit za dohvaćanje podataka iz baze podataka

Ad 5) DCL – koristi dvije naredbe:

GRANT - ova naredba dodjeljuje korisnicima privilegije pristupa bazi podataka.

REVOKE - naredba povlači korisnikove privilegije pristupa dane naredbom GRANT



Slika 18. Skupine SQL naredbi - izvor https://www.geeksforgeeks.org/sql-ddl-dql-dml-dcl-tcl-ommands/posjećeno 03.06.2022.

4.2.2 Funkcije u T-SQL

Kao i Excel i T-SQL omogućava korisniku sažimanje i grupiranje podataka te analizu podataka korištenjem različitih funkcija. Isto tako i spektar funkcija u T-SQL je vrlo širok. Ipak razlikujemo dvije osnovne skupine T-SQL funkcija:

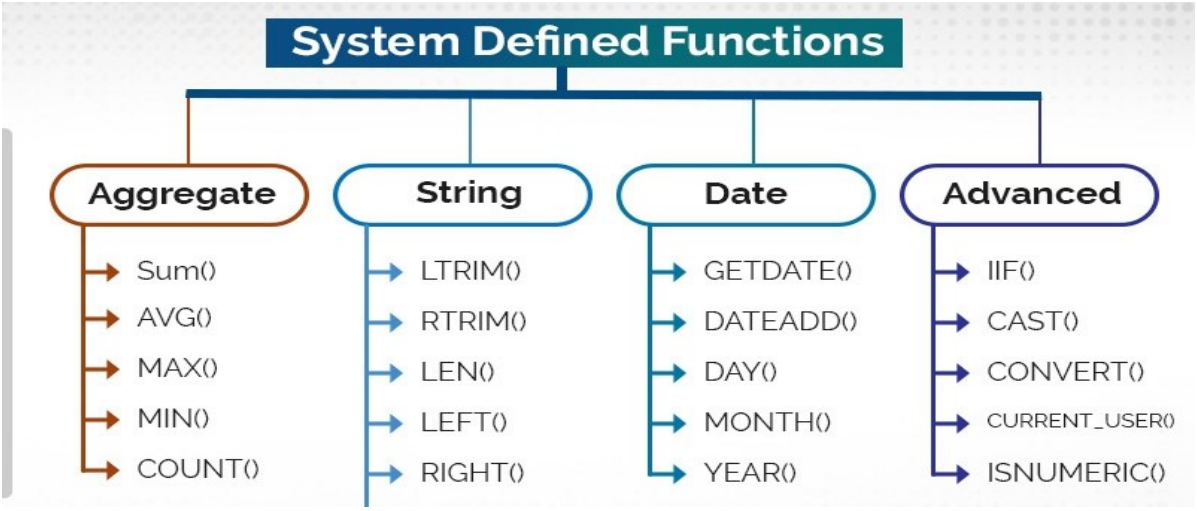
1. AGREGATNE koje još dijelimo na:

- Convenient agregatne funkcije (MIN, MAX, SUM, AVG, COUNT, COUNT_BIG)
- Statističke funkcije (VAR , VARP, STDEV, STDEVP)
- User defined functions
- Napredne funkcije za analizu podataka

2. SKALARNE koje još mogu biti:

- Numeričke,

- Funkcije dohvata i postavljanja datuma,
- String funkcije (funkcije niza znakova),
- Sistemske i
- Metadata funkcije



Slika 18. Skupine SQL naredbi - izvor https://www.geeksforgeeks.org/sql-ddl-dql-dml-dcl-tcl-ommands/posjećeno 03.06.2022.

4.2.3. Online obrada transakcija (OLTP) vs BI (Business Intelligence)

Petković (2019) konstatira da su se od početka sustavi relacijskih baza podataka koristili gotovo isključivo za prikupljanje primarnih poslovnih podataka, kao što su narudžbe i fakture, koristeći obradu temeljenu na transakcijama. Ovaj fokus na poslovne podatke ima svoje prednosti, ali i nedostatke. Jedna od prednosti je da su se loše performanse ranih sustava baza podataka dramatično poboljšale, do te mjere da danas mnogi sustavi baza podataka mogu izvršiti tisuće transakcija u sekundi (koristeći odgovarajući hardver). S druge strane, fokus na obradu transakcija zanemario je drugu prirodnu primjenu sustava baza podataka: njihovo korištenje za filtriranje i analizu potrebnih informacija iz svih postojećih podataka u organizaciji ili odjelu.

Sustave koji se temelje na obradi transakcija nazivamo sustavima za online obradu transakcija (OLTP)[5]. OLTP sustavi imaju neka važna svojstva, kao što su:

- Kratke transakcije, odnosno velika propusnost podataka
- Veliki broj korisnika

[5] OLTP -online transaction processing-D.Perković,"Microsoft SQL server 2019 A Begginers Guide 7th edition"

- Kontinuirane operacije čitanja i pisanja temeljene na malom broju redaka
- Podaci srednje veličine koji su pohranjeni u bazi podataka

Količina podataka pohranjenih u bazama podataka, posljednjih nekoliko godina, značajno se povećala, ali je još uvijek relativno mala u odnosu na količinu pohranjenu u skladištu podataka („dana warehouse“) za potrebe BI .
Petković (2019) definira poslovnu inteligencija (BI) kao proces integriranja podataka na razini poduzeća u jedinstvenu pohranu podataka iz koje krajnji korisnici mogu kreirati upite i izvješća za analizu postojećih podataka. Drugim riječima, cilj BI-a je čuvanje podataka kojima mogu pristupiti korisnici koji na temelju analize donose poslovne odluke. Ovi sustavi se često nazivaju analitički ili informativni sustavi jer pristupom podacima korisnici dobivaju potrebne informacije za donošenje boljih poslovnih odluka.

Razlike između OLTP sustava i BI su u njihovim osnovnim svojstvima, dok OLTP sustavi koriste kratke i učestale transakcije, a broj korisnika se kreće od nekoliko do nekoliko milijuna, BI ima potpuno drugačije odlike.

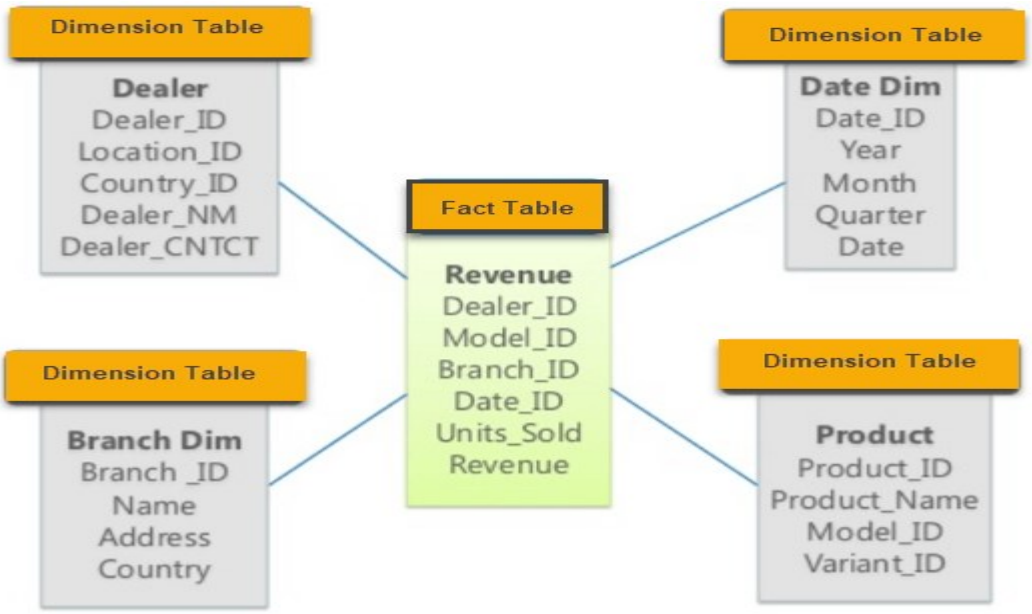
- Periodične operacije pisanja (učitavanja) s kompleksnim upitima na temelju velikog broja redaka
- Relativno mali ili poznati broj korisnika
- Ogromna količina podataka pohranjenih u bazi podataka

Još jedna bitna razlika između OLTP I BI je dostupnost podataka. Dok su podaci za OLTP sustave dostupni 24/7 u gotovo realnom vremenu, BI može koristiti podatke tek kada se isti učitaju u skladište podataka . OLTP sustavi sadrže trenutne podatke, dok BI prati povijesne podatke (uspoređivanje prodaje po različitim vremenskim razdobljima). To je glavni razlog velike količine podataka u skladištima podataka.
Skladišta podataka izvor su podataka za BI. Petković(2019) ih definira kao bazu podataka koja uključuje sve korporativne podatke i kojoj korisnici mogu jedinstveno pristupiti.
Organizacije i korporacije koriste različite sustave prikupljanja podataka. Prije spremanja u skladište podataka, podaci se moraju prikupiti, očistiti i konsolidirati. Konsolidacija podataka u skladištu je izuzetno važna jer informacije koje BI dohvati, temelj su za važne korporativne odluke. Konsolidacije se vrši tako što se ekstrahiraju podaci iz različitih izvora, takvi podaci se čiste, odnosno transformiraju, slijedi proces osiguravanja kvalitete podataka koji obično uključuje i suradnju s korisnicima.

Organizacije podataka u skladištu značajno se razlikuje od baza podataka. Dok je normalizacija poželjna kod baza podataka, skladišta rade na drugačiji način. Model koji koriste skladišta podataka naziva se dimenzijski model.

U dimenzijskom modeliranju, svaki pojedinačni model sastoji se od jedne tablice koja pohranjuje mjere, koja se naziva činjenična tablica („fact table“), i nekoliko drugih tablica koje opisuju dimenzije, koje se nazivaju dimenzijske tablice („dim_table“).

Činjenične tablice sadrže sve primarne ključeve dimenzijskih tablica i stupce atribute koje nazivamo mjere („measures“). To su atributi koje možemo analizirati po određenim dimenzijama. Najčešće korišten oblik organizacije tablica u skladištu podataka je star ili zvjezdana shema.



Slika 20.Zvjezdasta schema organizacije tablica – izvor [https://www.guru99.com/star-shema-in-data-warehouse-modeling.html -posjećeno 12.06.2022](https://www.guru99.com/star-shema-in-data-warehouse-modeling.html).

Jedan od glavnih razloga ovakve organizacije je smanjenje broja „join“ -ova odnosno povezivanja velikog broja tablica. Upiti kroz SELECT naredbu koju koristi BI vrlo su kompleksni i vremenski zahtjevni, kako u izradi tako i u izvođenju (često se radi o dohvatu nekoliko milijuna podataka) pa je pojednostavljenje upita imperativ za BI.

Još su dvije vrlo slične organizacijske sheme u skladištu podataka, a to su pahuljasta/snowflake) i galaksi shema.

Dizajn skladišta podataka vrlo je složen i dugotrajan posao, financijski neodrživ manjim organizacijama. Za njihove potrebe bolja opcija su „dana mart“ – ovi. To su manja skladišta podataka koja uključuju sve podatke na razini odjela i stoga korisnicima omogućuju pristup podacima koji se tiču samo jednog dijela njihove organizacije.

4.2.3.1. Analiza podataka u SSMS – praktičan primjer

Kako je već navedeno, najučinkovitiji način analize podatak ostvaruje se kroz skladišta podataka ili datamart-ove. Kako je ovaj rad vezan za program Poslovne Informatike na Veleučilištu u Šibeniku koji obuhvaća samo OLTP , analiza seta podataka radila se kroz transakcijsku bazu podataka.

Prije samog procesa analize bilo je potrebno izvršiti određene predradnje. Originalni set podataka Sales.xlsx preuzet je u formatu Excel radnog lista, to je zahtijevalo transformaciju podataka iz glavne tablice u nekoliko manjih relacijskih tablica svedenih na treću normalizacijsku formu bez redundancija. Tako kreirane tablice poslužit će za punjenje tablica u bazi podataka.

1. Query (upit) za kreiranje baze podataka s pripadajućim tablicama

```
CREATE DATABASE ANALYSIS
USE
ANALYSIS

CREATE TABLE TBLREGIONS
(REGION_NAME NVARCHAR(50),
REGION_ID CHAR(5) PRIMARY KEY NOT NULL,
)

CREATE TABLE TBLSTATES
(STATE_ID NVARCHAR(5) PRIMARY KEY NOT NULL,
STATE_NAME NVARCHAR (50),
REGION_ID CHAR(5) FOREIGN KEY REFERENCES TBLREGIONS(REGION_ID)
)

CREATE TABLE TBLCITIES
(CITY_ID INT PRIMARY KEY NOT NULL,
CITY_NAME NVARCHAR (50),
STATE_ID NVARCHAR(5) FOREIGN KEY REFERENCES TBLSTATES(STATE_ID) NOT NULL,
)

CREATE TABLE TBLSHIPMENTS
(SHIP_CODE INT PRIMARY KEY,
SHIP_MODE NVARCHAR (50)
)

CREATE TABLE TBLORDERS
(ORDER_NUMBER NVARCHAR (20),
ORDER_ID INT PRIMARY KEY,
ORDER_DATE DATE,
SHIP_DATE DATE,
SHIP_CODE INT FOREIGN KEY REFERENCES TBLSHIPMENTS(SHIP_CODE)
)

CREATE TABLE TBLSEGMENT
(SEGMENT_ID NVARCHAR (20) PRIMARY KEY,
SEGMENT_NAME NVARCHAR(30)
)
```

```
CREATE TABLE TBLCUSTOMERS
(CUSTOMER_ID INT PRIMARY KEY NOT NULL,
CUSTOMER_NUMBER NVARCHAR(20),
CUST_FIRST_NAME NVARCHAR(30),
CUST_LAST_NAME NVARCHAR (50),
SEGMENT_ID NVARCHAR (20) FOREIGN KEY REFERENCES TBLSEGMENT(SEGMENT_ID),
CITY_ID INT FOREIGN KEY REFERENCES TBLCITIES(CITY_ID) )

CREATE TABLE TBLCATEGORIES
(CATEGORY_NAME NVARCHAR (20),
CATEGORY_ID NVARCHAR(5) PRIMARY KEY
)

CREATE TABLE TBLSUBCATEGORIES
(SUBCATEGORY_NAME NVARCHAR(30),
SUBCAT_ID NVARCHAR(10) PRIMARY KEY,
CATEGORY_ID NVARCHAR(5) FOREIGN KEY REFERENCES TBLCATEGORIES(CATEGORY_ID)
)

CREATE TABLE TBLPRODUCTS
(PRODUCT_ID NVARCHAR (50) PRIMARY KEY,
PRODUCT_NAME NVARCHAR(100),
SUBCAT_ID NVARCHAR(10) FOREIGN KEY REFERENCES TBLSUBCATEGORIES(SUBCAT_ID),

)

CREATE TABLE TBLSALES
(SALES_ID INT PRIMARY KEY,
ORDER_ID INT FOREIGN KEY (ORDER_ID) REFERENCES TBLORDERS(ORDER_ID),
ORDER_DATE DATE,
SHIP_DATE DATE,
CUSTOMER_ID INT FOREIGN KEY REFERENCES TBLCUSTOMERS(CUSTOMER_ID),
PRODUCT_ID NVARCHAR (50) FOREIGN KEY REFERENCES TBLPRODUCTS(PRODUCT_ID),
SALES DECIMAL (12,4),
QUANTITY INT,
DISCOUNT DECIMAL (7,3),

)
```

Na SSMS je kreirana nova baza podataka po nazivom **Analysis** i 11 novih tablica:

tblCategories

tblCities

tblCustomers

tblOrders

tblProducts

tblRegions

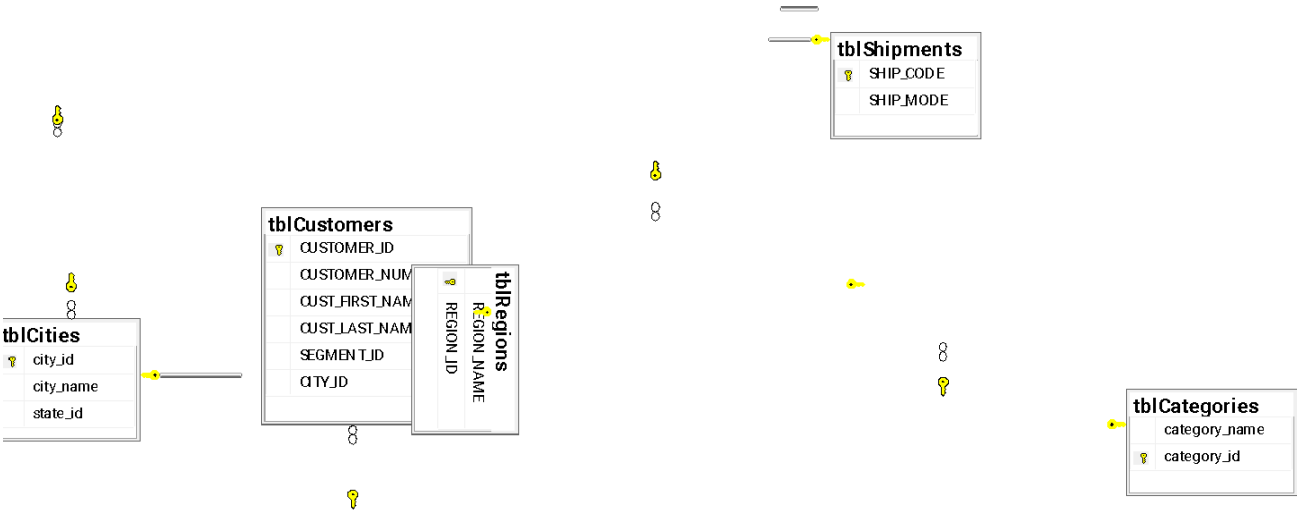
tblSales

tblSegment

tblShipments

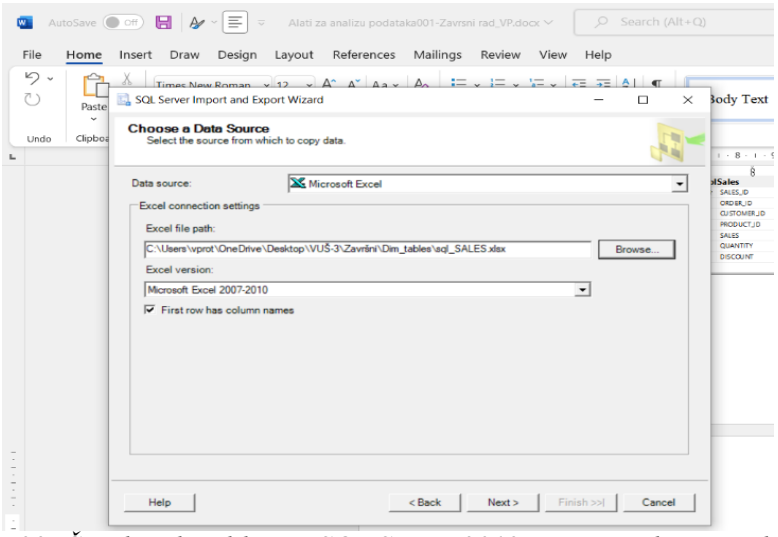
tblStates

tblSubCategories



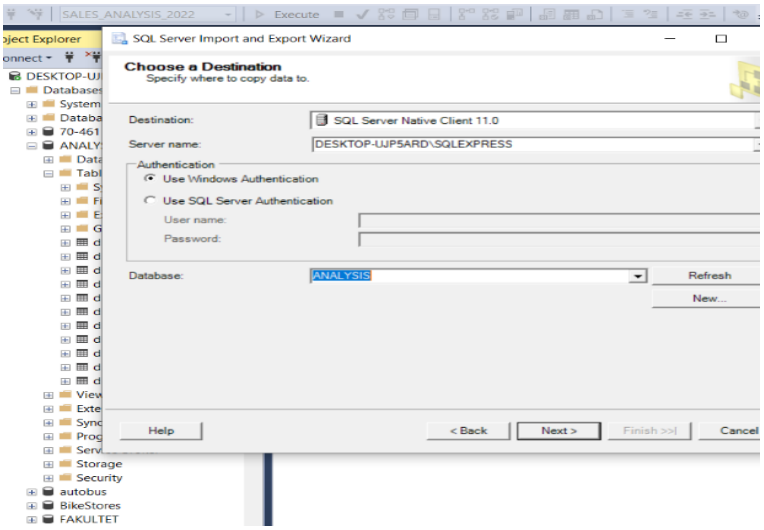
Slika 21. Baza podataka Analysis

Set podataka koji je korišten u radu sadrži 9994 redaka, pa smo kao alat za punjenje tablica podacima koristili SQL Server 2019 Import and Export Data (64-bit) aplikaciju. Čarobnjak (eng. wizard) vas vodi kroz proces. Potrebno je odabrati datoteku koja će poslužiti kao izvor podataka.



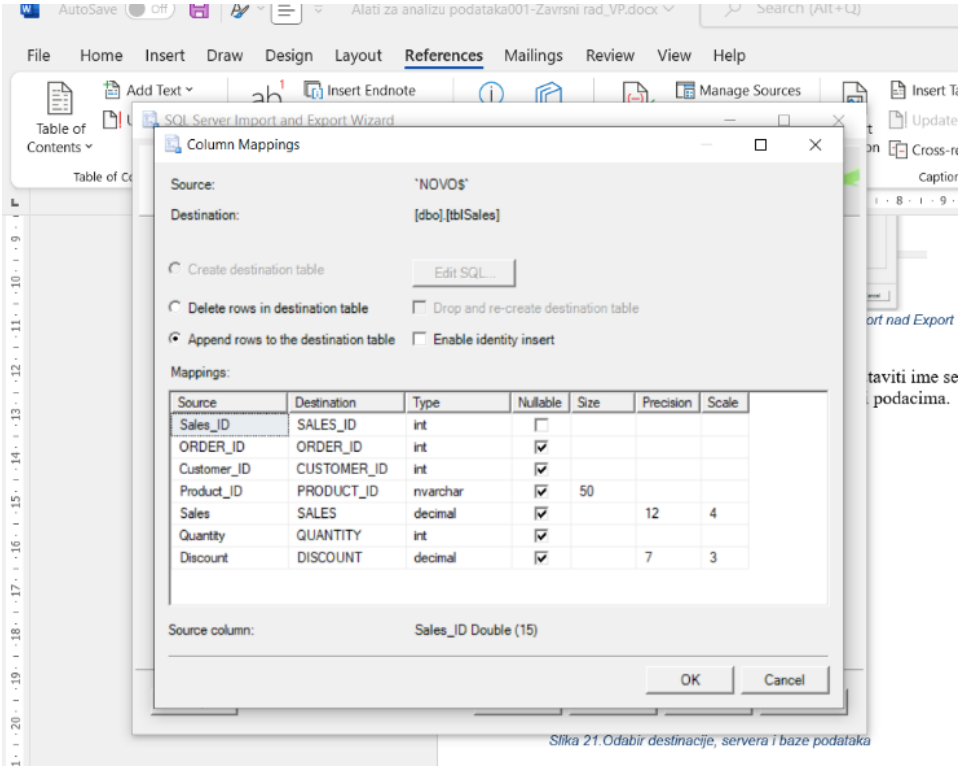
Slika 22. Čarobnjak aplikacije SQL Server 2019 Import and Export dana

Zatim je potrebno odabrati destinaciju, postaviti ime servera te odabrati bazu podataka u kojoj se nalazi tablica koju želimo napuniti podacima.



Slika 23.Odabir destinacije, servera i baze podataka

U sljedećem izborniku otvara se mogućnost povezivanja izvorne tablice s odredišnom tablicom. Nakon toga moguće je urediti mapiranje i pregledati odgovaraju li nazivi stupaca i tipovi podataka odredišnim tablicama.

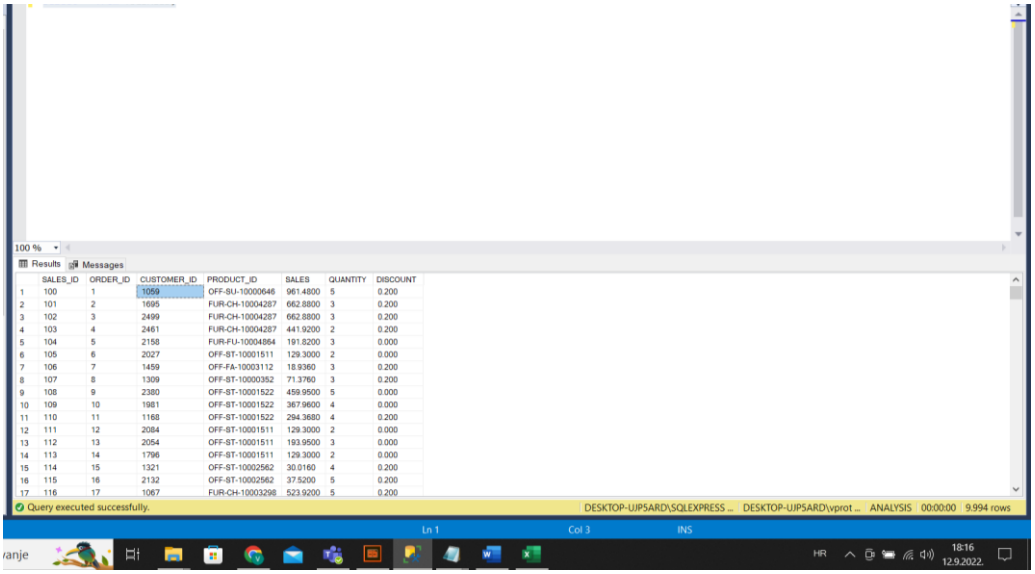


Slika 24.Uređivanje mapiranja prije punjenja podataka u odredišnu tablicu

Po završetku uređivanja mapiranja potrebno je potvrditi odabir i odabirom opcije slijedeće podaci će se učitati u odgovarajuću tablicu.

Provjeru učitanih podataka vršimo pomoću naredbe SELECT.

```
--select * from TBLSALES;
```



ili

SELECT COUNT(*) as ukupan_broj_redaka FROM TBLSALES
ako želimo biti sigurni da su se učitali svi redci.

ukupan_broj_redaka
9994

Analiza podataka neovisno o alatu koji koristimo, započinje postavljanjem pitanja na koja ona treba odgovoriti . U ovom dijelu izrađeni su upiti koji donose podatke o navikama kupaca, ukupno ostvarenoj prodaji u pojedinom segmentu te izradu „view“-a za jednostavniju analizu kompleksnih upita.

Odgovore koje tražimo u Upitu 1 su:

1. koliko imamo stalnih kupaca
2. a koliko onih koji su obavili samo jednu kupnju
3. koliki je njihov omjer u ukupnom broju kupaca

2. Upit (Query) – navike kupaca

Kreiranje upita koji vraća broj kupaca koji su ostvarili jednu kupnju i broj kupaca koji se vraćaju odnosno koji su os
ostvarili više od jedne kupnje. Upit je kreiran uz korištenje CTE (common table expression[6]). Rezultat upita moguće je izvesti u CSV ili neki drugi oblik datoteke.

[6] Common table expression je privremeni imenovani skup rezultata na koji se možemo pozivati unutar naredbe SELECT, INSERT, UPDATE ili DELETE. CTE se također može koristiti u View-u.

```sql
with buyers as
(select CUSTOMER_ID,ROW_NUMBER()OVER (PARTITION BY CUSTOMER_ID ORDER BY CUSTOMER_ID)
AS ROWNUM
FROM tblSales S),
REPEAT_BUYERS AS
(SELECT DISTINCT CUSTOMER_ID
 FROM buyers
 where ROWNUM >1
)
, ONE_PURCHASE as (
select distinct
         CUSTOMER_ID
FROM buyers
where  rownum = 1
)
select
       case
              when r.CUSTOMER_ID IS NOT NULL
              then 'repeat'
              else 'new'
              end
              as repeat_new,
              count(*) as NUMBER_OF_CUSTOMER,
              (      select count(distinct CUSTOMER_ID)
                     from buyers
                     ) as TOTAL_CUSTOMERS,
                     FORMAT(cast(count(*) as decimal(18,2))/cast((select count(distinct
                    CUSTOMER_ID) from buyers) as decimal(18,2)), 'P') as REPEAT_RATE

from ONE_PURCHASE O
left join  repeat_buyers r on O.CUSTOMER_ID = r.CUSTOMER_ID
group by
       case
              when r.CUSTOMER_ID IS NOT NULL
              then 'repeat'
              else 'new'
              end;
```

Tablica 2.

repeat_new	BROJ_KUPACA	UKUPAN_BROJ_KUPACA	PONOVLJENA _KUPNJA_%
repeat	788	793	99.37%
new	5	793	0.63%

Navike kupaca

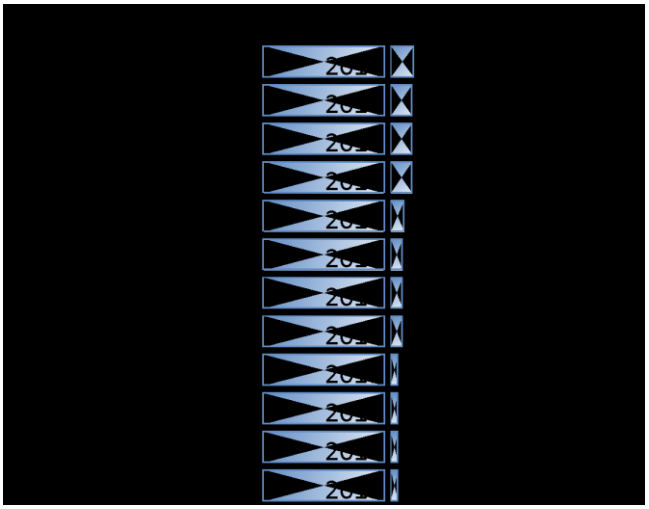
Analiza podataka pokazuje da od ukupnog broja kupaca , 99.37% su stalni kupci, odnosno kupci koji su obavili više od jedne transakcije, dok je udio novih kupaca, odnosno kupaca koji su obavili samo jednu transakciju manji od 1%,

Slijedeći upit dohvaća podatke o ukupnom broju kupaca po godinama i segmentu. Na taj način dobit ćemo uvid u kretanje kupaca, odnosno porast ili smanjenje broja kupaca u određenom segmentu. Na temelju dobivenih rezultata, odjel upravljanja odnosima s kupcima može razvijati nove strategije.

3. Query (Upit) Kretanje kupaca

```
--ANALIZA KUPACA --UPIT KOJI DOHVAĆA UKUPAN BROJ KUPACA ZA POJEDINI SEGMENT I GODINU
SELECT S.SEGMENT_NAME,C.SEGMENT_ID, DATEPART(YEAR,ORDER_DATE) AS
'YEAR',count(distinct(f.customer_id)) as Number_of_Customers
FROM
tblCustomers C
JOIN
tblSegment S
ON
S.SEGMENT_ID=C.SEGMENT_ID
JOIN
tblSales F
ON
C.CUSTOMER_ID = F.CUSTOMER_ID
JOIN
tblOrders O
ON
F.ORDER_ID=O.ORDER_ID
--WHERE DATEPART(YEAR,ORDER_DATE) = '2015'
GROUP BY S.SEGMENT_NAME,C.SEGMENT_ID, DATEPART(YEAR,ORDER_DATE)
 ORDER by Number_of_Customers DESC;
```

Tablica 3.



Kretanje kupaca

Događa se da pri analizi velikih i kompleksnih tablica trebamo samo određeni set podataka koji analiziramo detaljnije. Kako ne bismo više puta zadavali isti ili sličan upit bazi podataka, kreiramo View(pogled). Podatke iz view-a možemo dalje analizirati neovisno o tablici iz koje je kreiran. Ipak treba imati na umu da svakim pozivanjem view dohvaća nove (update) podatke iz baze. Ako želimo analizirati isti set podataka dulje vremensko razdoblje bolja opcija je kreiranje materijaliziranog viewa koji će dohvaćati nove podatke samo kada ga osvježimo(refresh)**.** Izradili smo view koji dohvaća samo prvih 10 proizvoda radi jednostavnije prezentacije rezultata. Ako zanemarimo naredbu TOP, možemo kreirati view koji će dohvaćati samo proizvode za određenu godinu, koje dalje možemo analizirati bez kompleksnih upita na bazu.

4. Query – izrada View

```
--TOP 10 proizvoda za danu godinu
CREATE OR ALTER VIEW Products_2018 AS
SELECT  top 10 PRODUCT_NAME, SUBCAT_ID, COUNT(F.PRODUCT_ID) AS NUMBER_OF_PRODUCTS ,
DATEPART(YEAR,ORDER_DATE) AS 'YEAR',
SUM(F.SALES) AS TOTAL_SALES
FROM
tblSales F
JOIN tblProducts P
ON
F.PRODUCT_ID =P.PRODUCT_ID
JOIN
tblOrders O
ON
F.ORDER_ID=O.ORDER_ID
WHERE DATEPART(YEAR,ORDER_DATE)  = '2018'
GROUP BY PRODUCT_NAME,DATEPART(YEAR,ORDER_DATE),SUBCAT_ID
GO

--dohvat podataka vršimo kao i kod obične tablice
 SELECT * FROM Products_2018
```

Tablica 4.

View products_2018

5. ZAKLJUČAK

U današnje doba digitalne transformacije i velikih podataka, analiza podataka postala je neizostavan dio poslovanja , ali i života. Sve češće čujemo pojam „Data is a new oil“ (podaci su nova nafta), što bi značilo da uvidi koje dobijemo iz kvalitetnih podataka, kao što su mogućnost predviđanja ponašanje kupca ili anticipiranje tržišnih trendova ili otkrivanje složenih obrazaca i korelacija, prije konkurencije, osiguravaju kompanijama prednost koja može biti presudna u njihovom poslovanju. Danas se u analizi podataka koriste i složeniji alati od onih koje ovaj rad obrađuje, a zasnivaju se na tehnikama rudarenja podataka, umjetnoj inteligenciji, strojnom učenju itd. Ipak neki alati zadržavaju svoju poziciju kao neizostavan dio analize podataka. Excel (uz svoje „add-ins“opcije) jedan je od takvih alata, kompleksan u strukturi, a relativno jednostavan za korištenje. Proširenja Excela kao što su Rješavač, Data Analysis tools, Power Query i mnogi drugi, napredna su rješenja koja izvode kompleksne statističke analize u vrlo kratkom vremenu. Ali kao i svaki alat, Excel ima svoja ograničenja. Odlično se ponaša kod manjih setova podataka, ali u okruženju od nekoliko stotina tisuća ili milijuna podataka, obrada traje dugo i tu alat gubi svoje prednosti. SSMS (SQL Server Management Studio) je alat koji se odlično ponaša kod obrade velikih skupova podataka. T-SQL programski je jezik koji koristi SSMS (to je inačica SQL programskog jezika). Osim mogućnosti kreiranja baza podataka, a uz dodatne komponente i skladišta podataka, ovaj alat omogućava upravljanje dostupnošću podataka, tako korisnici mogu pristupiti točno određenim podacima, a ne svim podacima. Baze podataka ali i skladišta podataka temelju su onog što danas nazivamo analiza podataka. Velike korporacije podatke prikupljaju u pomno dizajnirana skladištima podataka u kojima su pohranjeni povijesni podaci, prostorno i vremenski kontekstualizirani i iz kojih korisnici dohvaćaju različite izvještaje. Dizajn skladišta je vremenski i financijski vrlo zahtjevan postupak koji u prosjeku traje dvije godine i stoji oko 5 milijuna dolara (Petković 2019). Male organizacije si ne mogu priuštiti takve investicije, ali njima su na raspolaganju „data mart“-ovi (mala skladišta podataka) vezana isključivo za specifičan odjel u organizaciji. Nisu samo velike korporacije te koje koriste analizu podataka, vlade, udruge, ali i kućanstva, prikupljaju i analiziraju podatke za donošenje bolji i kvalitetnijih odluka Skladišta podataka, područja su gdje nas očekuju velike promjene, kako u optimizaciji prostora za pohranu, tako i u brzini dohvata podataka i kreiranja izvještaja. Neovisno u kakvom obliku su pohranjeni podaci za analizu, njihova dostupnost i sigurnost bit će imperativ analize podataka u budućnosti.

LITERATURA

1. Linoff S.Gordon „Data analysis using Excel and SQL“, John Wiley & Sons, Inc,2021.
2. McFedries Paul „ Excel data analysis for dummies 7th edition“
3. Wayne L.Winston „Microsoft Excel Data Analysis and Business Modeling (Office 2021 and Microsoft 365)7th Edition, Microsoft 2021.
4. Guerrero Hector,“ Excel Data Analysis_ Modeling and Simulation“Springer (2019)
5. Rees MuchaelPrinciples of Financial Modelling_ Model Design and Best Practices Using Excel and VBA,Wiley
6. Petković Dušan,“Microsoft SQL Server 2019 A beginner's guide“ McGraw Hill 2020.
7. https://www.geeksforgeeks.org/sql-ddl-dql-dml-dcl-tcl-ommands/posjećeno 03.06.2022.
8. https://www.guru99.com/star-schema-in-data-warehouse-modeling.html -posjećeno 12.06.2022.
9. https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver16 posjećeno 29.05.2022.
10. https://www.techtarget.com/searchdatamanagement/definition/SQL posjećeno 03.06.2022,
11. www.simplelearn.com – *posjećeno 12.05.2022.*

PRILOZI

1.1 Popis slika

Slika 1. ETL proces u Power Query 6

Slika 2. Excel tablica s 9995 redaka 9

Slika 3. Učitavanje tablice sales_data.csv u PQ 9

Slika 4. Postavljanje željenog formata datuma – izvor autor 10

Slika 5. Dodavanje nove kolone uz korištenje formula za izračun iznosa 10

Slika 6. Stvaranje veza na tablicu 11

Slika 7. Povezivanje tablica putem MERGE opcije 11

Slika 8. Tablica sales 14

Slika 9. Izvještaj prodaje - Sales Report -izvor autor 15

Slika 10. Izvještaj analiza kupaca- Customer report -izvor autor 16

Slika 11.Izvještaj analize proizvoda – Products report -izvor autor 16
Slika 12. Polja pivot grafa - izvor "Excel data analysis for Dummies" 18
Slika 13 Kreiranje pivot tablice 19
Slika 14.Usporedne pivot tablice 19
Slika 16. Izvještaj o godišnjoj prodaji proizvoda 20
Slika 17. Grafikon mapa -Unique customer report 21
Slika 18. Skupine SQL naredbi 25
Slika 19. Zvjezdasta shema organizacije tablica 30
Slika 20. Baza podataka Analysis 30
Slika 21. Uređivanje mapiranja prije punjenja podataka u odredišnu tablicu 30
Slika 22. Čarobnjak aplikacije SQL Server 2019 Import and Export dana 31
Slika 23. Odabir destinacije, servera i baze podataka 31
Slika 24. Uređivanje mapiranja prije punjenja podataka u odredišnu tablicu 32

1.2 Popis tablica

Tablica 1. Pojmovnik tablice Sales 7
Tablica 2. Navike kupaca 33
Tablica 3. Kretanje kupaca 34
Tablica 4. View products_2018 35