

# Univerzitet Singidunum

Fakultet za informatiku i računarstvo

## Obrada podataka o prodaji sa ecommerce sajta Amazon.com

Drugi kolokvijum iz predmeta Veliki skupovi podataka

Profesor: Dalibor Radovanović

Student: Nikola Nenadović

Broj indeksa: 2023205687

# I. Uvod

## Tableau

Tableau je softverski alat namenjen analizi i vizualizaciji podataka sa fokusom na poslovnu inteligenciju(eng. Business Intelligence). Tableau se prvi put pojavljuje 1999. godine kao softver za vizualizaciju podataka pod imenom Polaris, razvijen od strane Patricka Hanrahana i Chrisa Stoltea, kao rezultat programa Accelerated Strategic Computing Initiative(ASCI), podržanog od strane Ministarstva energetike Sjedinjenih Američkih Država. Kompaniju Tableau su januara 2003. godine formalno osnovali Christian Chabot, Stolte i Hanrahan na univerzitetu Stanford, a sedište firme već sledećeg meseca premestili u Seattle, Washington, gde se i danas nalazi. Od samog osnivanja, pa do današnjeg dana, Tableau se izuzetno brzo razvija i uspešno posluje u kompetitivnom i inovativnom poslovnom okruženju, što je rezultiralo prodajom firme kompaniji Salesforce, 10. Juna 2019. godine. Vrednost transakcije iznosila je 15,9 milijardi dolara.

Ključne prednosti Tablea su:

- Intuitivan korisnički interfejs: Prijemčivi drag-and-drop način korišćenja omogućava analizu i vizualizaciju podataka bez dodatnog programiranja
- Lakoća predstavljanja podataka: Tableau transformiše sirove podatke u tabele, grafove i druge vrste grafičke reprezentacije. Tableau se može povezati sa tabelama, bazama podataka, cloud i Big Data platformama i vršiti analizu i prezentaciju podataka na jasan i razumljiv način
- Integracija sa različitim izvorima podataka: Tableau se povezuje sa velikim brojem popularnih servisa kao što su Microsoft Excel, SQL baze podataka, Google Analytics, Salesforce, Amazon Web Services(AWS),
- Dostupnost resursa za obuku i podrška zajednice: Tableau je izgradio pouzdanu mrežu korisnika, programera i analitičara koja razvija materijale za učenje, sertifikaciju i unapređenje znanja i veština analize i reprezentacije podataka

Tableau softver je znatno pojednostavio proces interpretacije i grafičkog predstavljanja podataka i tako približio oblast analize podataka ogromnom broju potencijalnih korisnika. Mnoge organizacije iz raznolikih industrijskih oblasti koriste Tableau za kreiranje izveštaja, praćenje i unapređenje efikasnosti, i kao putokaz pri donošenju važnih poslovnih odluka. Najpopularniju softverski proizvodi Tableau kompanije su Tableau Desktop, Tableau Server, Tableau Public, Tableau Online i Tableau Prep.

## II. Pronalazak odgovarajućeg dataseta

Podaci koje ćemo analizirati u projektu su iz dataseta *Amazon Sales*, autora Ananda Shawa, nalaze se na sajtu Kaggle.com i omogućena je njihova javna upotreba. Izvor podataka su sakupljene informacije o prodaji fizičkih proizvoda sa sajta Amazon.com koje su se odvijale na tržištu Sjedinjenih Američkih Država, uglavnom u periodu od 2011. do 2014. godine. Dataset se sastoji iz 19 kolona i dve tabele, *Sales* i *Customers*. Kolone tabele *Sales* su:

- *OrderID*: Niz od 14 karaktera, sastavljen od slova i brojeva, koji jedinstveno identifikuje svaku transakciju u tabeli
- *Order Date*: Datum naručivanja proizvoda u formatu *dd.m.yyyy*
- *Ship Date*: Datum isporuke proizvoda, takođe u formatu *dd.mm.yyyy*
- *EmailID*: Email adresa kupca koja istovremeno služi i kao korisničko ime
- *Country, State, City*: Kolone koje tekstualnim tipom podataka označavaju zemlju, državu i grad iz kojeg se vrši porudžbina
- *Category*: Kategorije naručenih proizvoda
- *Product Name*: Puno ime proizvoda koji je predmet transakcije
- *Sales*: Učinak prodaje, izražen u decimalnim brojevima
- *Quantity*: Količina prodatog proizvoda
- *Profit*: Profit po svakom proizvodu, izražen u decimalnim brojevima
- *Order year, Ship year*: Godina naručivanja i godina isporuke

Tabela *Customers* sadrži sledeće kolone:

*CustomerID*: Jedinstveni identifikacioni broj kupca

*CustomerName*: Ime i prezime kupca

*Gender*: Pol kupca

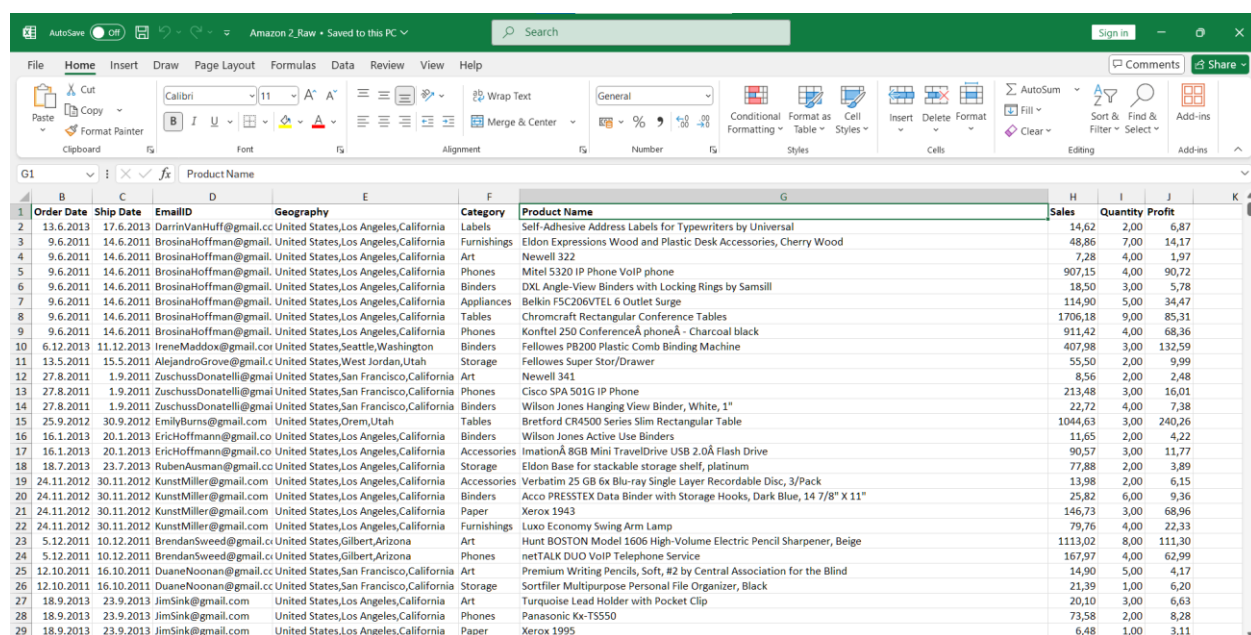
*OrderNumber*: Niz od 14 karaktera, slova i brojeva koji jedinstveno označava svaku narudžbinu

*PaymentMethod*: Načini plaćanja predstavljeni tekstualnim tipom podataka

Dataset je u .xlsx formatu, sastoji se od ukupno 17 kolona i 3204 reda, sadrži minimalan broj NULL vrednosti te nisu potrebne velike aktivnosti u vezi sa čišćenjem podataka, pa odmah možemo započeti analizu i vizualizaciju.

### III. Ubacivanje i čišćenje podataka u Tableau softveru

Iako su podaci u datasetu u odličnom stanju, nakon kraćeg razmatranja rešio sam da pristupim postupcima preuređenja i čišćenje podataka. Nakon prvobitnog pregleda dataseta, analize kolona i redova, provere koliko redova sadrži NULL vrednosti, obratio sam pažnju na kolonu Geography, koja sadrži lokaciju naručioca proizvoda, predstavljenu u formi Zemlja, Grad, Država (npr. United States, Gilbert, Arizona). Izvršio sam podelu kolone Geography u tri nove kolone: Country, City, State. Delimiter prilikom razdvajanja kolone je karakter „ „ , „ , a kolonu Geography sam uklonio.



Order Date	Ship Date	EmailID	Geography	Category	Product Name	Sales	Quantity	Profit
13.6.2013	17.6.2013	DarrinVaniHuff@gmail.cc	United States, Los Angeles, California	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14,62	2,00	6,87
9.6.2011	14.6.2011	BrosinaHoffman@gmail.	United States, Los Angeles, California	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48,86	7,00	14,17
9.6.2011	14.6.2011	BrosinaHoffman@gmail.	United States, Los Angeles, California	Art	Newell 322	7,28	4,00	1,97
9.6.2011	14.6.2011	BrosinaHoffman@gmail.	United States, Los Angeles, California	Phones	Mitel 5320 IP Phone VoIP phone	907,15	4,00	90,72
9.6.2011	14.6.2011	BrosinaHoffman@gmail.	United States, Los Angeles, California	Binders	DXL Angle-View Binders with Locking Rings by Samsill	18,50	3,00	5,78
9.6.2011	14.6.2011	BrosinaHoffman@gmail.	United States, Los Angeles, California	Appliances	Belkin F5C206VTEL 6 Outlet Surge	114,90	5,00	34,47
9.6.2011	14.6.2011	BrosinaHoffman@gmail.	United States, Los Angeles, California	Tables	Chromcraft Rectangular Conference Tables	1706,18	9,00	85,31
9.6.2011	14.6.2011	BrosinaHoffman@gmail.	United States, Los Angeles, California	Phones	Konftel 250 Conference phone - Charcoal black	911,42	4,00	68,36
6.12.2013	11.12.2013	IreneMaddox@gmail.com	United States, Seattle, Washington	Binders	Fellowes PB200 Plastic Comb Binding Machine	407,98	3,00	132,59
13.5.2011	15.5.2011	AlejandroGrove@gmail.c	United States, West Jordan, Utah	Storage	Fellowes Super Stor/Drawer	55,50	2,00	9,99
27.8.2011	1.9.2011	ZuschussDonatelli@gmail	United States, San Francisco, California	Art	Newell 341	8,56	2,00	2,48
27.8.2011	1.9.2011	ZuschussDonatelli@gmail	United States, San Francisco, California	Phones	Cisco SPA 501G IP Phone	213,48	3,00	16,01
27.8.2011	1.9.2011	ZuschussDonatelli@gmail	United States, San Francisco, California	Binders	Wilson Jones Hanging View Binder, White, 1"	22,72	4,00	7,38
25.9.2012	30.9.2012	EmilyBurns@gmail.com	United States, Orem, Utah	Tables	Bretford CR4500 Series Slim Rectangular Table	1044,63	3,00	240,26
16.1.2013	20.1.2013	EricHoffmann@gmail.co	United States, Los Angeles, California	Binders	Wilson Jones Active Use Binders	11,65	2,00	4,22
16.1.2013	20.1.2013	EricHoffmann@gmail.co	United States, Los Angeles, California	Accessories	Imation 8GB Mini TravelDrive USB 2.0 Flash Drive	90,57	3,00	11,77
18.7.2013	23.7.2013	RubenAusman@gmail.co	United States, Los Angeles, California	Storage	Eldon Base for stackable storage shelf, platinum	77,88	2,00	3,89
24.11.2012	30.11.2012	KunstMiller@gmail.com	United States, Los Angeles, California	Accessories	Verbatim 25 GB 6x Blu-ray Single Layer Recordable Disc, 3/Pack	13,98	2,00	6,15
24.11.2012	30.11.2012	KunstMiller@gmail.com	United States, Los Angeles, California	Binders	Acco PRESTEX Data Binder with Storage Hooks, Dark Blue, 14 7/8" X 11"	25,82	6,00	9,36
24.11.2012	30.11.2012	KunstMiller@gmail.com	United States, Los Angeles, California	Paper	Xerox 1943	146,73	3,00	68,96
24.11.2012	30.11.2012	KunstMiller@gmail.com	United States, Los Angeles, California	Furnishings	Luxo Economy Swing Arm Lamp	79,76	4,00	22,33
5.12.2011	10.12.2011	BrendanSweed@gmail.c	United States, Gilbert, Arizona	Art	Hunt BOSTON Model 1606 High-Volume Electric Pencil Sharpener, Beige	1113,02	8,00	111,30
5.12.2011	10.12.2011	BrendanSweed@gmail.c	United States, Gilbert, Arizona	Phones	netTALK DUO VoIP Telephone Service	167,97	4,00	62,99
12.10.2011	16.10.2011	DuaneNoonan@gmail.cc	United States, San Francisco, California	Art	Premium Writing Pencils, Soft, #2 by Central Association for the Blind	14,90	5,00	4,17
12.10.2011	16.10.2011	DuaneNoonan@gmail.cc	United States, San Francisco, California	Storage	Sortiflier Multipurpose Personal File Organizer, Black	21,39	1,00	6,20
18.9.2013	23.9.2013	JimSink@gmail.com	United States, Los Angeles, California	Art	Turquoise Lead Holder with Pocket Clip	20,10	3,00	6,63
18.9.2013	23.9.2013	JimSink@gmail.com	United States, Los Angeles, California	Phones	Panasonic Kx-T5550	73,58	2,00	8,28
18.9.2013	23.9.2013	JimSink@gmail.com	United States, Los Angeles, California	Paper	Xerox 1995	6,48	1,00	3,11

Slika 1 – Početno stanje Amazon Sales dataseta

Takođe, kreirao sam dve nove kolone: Order Year i Ship Year. Vrednosti redova tih kolona su godina naručivanja proizvoda i godina isporuke proizvoda, a vrednosti podataka „izvučene“ su iz kolona Order Date i Ship Date. Pretvaranje postojećih i dodavanje novih kolona izvršio sam zbog efikasnije analize podataka i bolje preglednosti prilikom procesa vizualizacije. Smatram da prilikom preuređenja kolona dataseta nisam narušio integritet podataka, zato što podaci nisu bitno menjani, već je izvršena promena njihove forme radi efikasnijeg i kvalitetnijeg rada sa njima u budućnosti.

	D	E	F	G	H	I	J	K	L	M	N
	EmailID	Country	State	City	Category	Product Name	Sales	Quantity	Profit	Order Year	Ship Date
1	DarrinVanHuff@gmail.com	United States	California	Los Angeles	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14,62	2	6,8714	2013	
2	BrosinaHoffman@gmail.com	United States	California	Los Angeles	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48,86	7	14,1694	2011	
3	BrosinaHoffman@gmail.com	United States	California	Los Angeles	Art	Newell 322	7,28	4	1,9656	2011	
4	BrosinaHoffman@gmail.com	United States	California	Los Angeles	Phones	Mitel 5320 IP Phone VoIP phone	907,152	4	90,7152	2011	
5	BrosinaHoffman@gmail.com	United States	California	Los Angeles	Binders	DXL Angle-View Binders with Locking Rings by Samsill	18,504	3	5,7825	2011	
6	BrosinaHoffman@gmail.com	United States	California	Los Angeles	Appliances	Belkin F5C206VTEL 6 Outlet Surge	114,9	5	34,47	2011	
7	BrosinaHoffman@gmail.com	United States	California	Los Angeles	Tables	Chromcraft Rectangular Conference Tables	1706,184	9	85,3092	2011	
8	BrosinaHoffman@gmail.com	United States	California	Los Angeles	Phones	Konftel 250 ConferenceA phoneA - Charcoal black	911,424	4	68,3568	2011	
9	IreneMaddox@gmail.com	United States	Washington	Seattle	Binders	Fellowes PB200 Plastic Comb Binding Machine	407,976	3	132,5922	2013	
10	AlejandroGrove@gmail.com	United States	Utah	West Jordan	Storage	Fellowes Super Stor/Drawer	55,5	2	9,99	2011	
11	ZuschussDonatelli@gmail.com	United States	California	San Francisco	Art	Newell 341	8,56	2	2,4824	2011	
12	ZuschussDonatelli@gmail.com	United States	California	San Francisco	Phones	Cisco SPA 501G IP Phone	213,48	3	16,011	2011	
13	ZuschussDonatelli@gmail.com	United States	California	San Francisco	Binders	Wilson Jones Hanging View Binder, White, 1"	22,72	4	7,384	2011	
14	EmilyBurns@gmail.com	United States	Utah	Orem	Tables	Bretford CR4500 Series Slim Rectangular Table	1044,63	3	240,2649	2012	
15	ErichHoffmann@gmail.com	United States	California	Los Angeles	Binders	Wilson Jones Active Use Binders	11,648	2	4,2224	2013	
16	ErichHoffmann@gmail.com	United States	California	Los Angeles	Accessories	Imation® B08 Mini TravelDrive USB 2.0A Flash Drive	90,57	3	11,7741	2013	
17	RubenKusman@gmail.com	United States	California	Los Angeles	Storage	Eldon Base for stackable storage shell, platinum	77,88	2	3,884	2013	
18	KunstMiller@gmail.com	United States	California	Los Angeles	Accessories	Verbatim 25 GB 6x Blu-ray Single Layer Recordable Disc, 3/Pack	13,98	2	6,1512	2012	
19	KunstMiller@gmail.com	United States	California	Los Angeles	Binders	Acco PRESSTEX Data Binder with Storage Hooks, Dark Blue, 14 7/8" X 11"	25,824	6	9,3612	2012	
20	KunstMiller@gmail.com	United States	California	Los Angeles	Paper	Xerox 1943	146,73	3	68,9631	2012	
21	KunstMiller@gmail.com	United States	California	Los Angeles	Furnishings	Luxo Economy Swing Arm Lamp	79,76	4	22,3328	2012	
22	BrendanSweed@gmail.com	United States	Arizona	Gilbert	Art	Hunt BOSTON Model 1606 High-Volume Electric Pencil Sharpener, Beige	1113,024	8	111,3024	2011	
23	BrendanSweed@gmail.com	United States	Arizona	Gilbert	Phones	netTALK DUO VoIP Telephone Service	167,968	4	62,988	2011	
24	DuaneNoonan@gmail.com	United States	California	San Francisco	Art	Premium Writing Pencils, Soft, #2 by Central Association for the Blind	14,9	5	4,172	2011	
25	DuaneNoonan@gmail.com	United States	California	San Francisco	Storage	Sortliff Multipurpose Personal File Organizer, Black	21,39	1	6,2031	2011	
26	JimSink@gmail.com	United States	California	Los Angeles	Art	Turquoise Lead Holder with Pocket Clip	20,1	3	6,633	2013	
27	JimSink@gmail.com	United States	California	Los Angeles	Phones	Panasonic Kx-TS550	73,584	2	8,2782	2013	

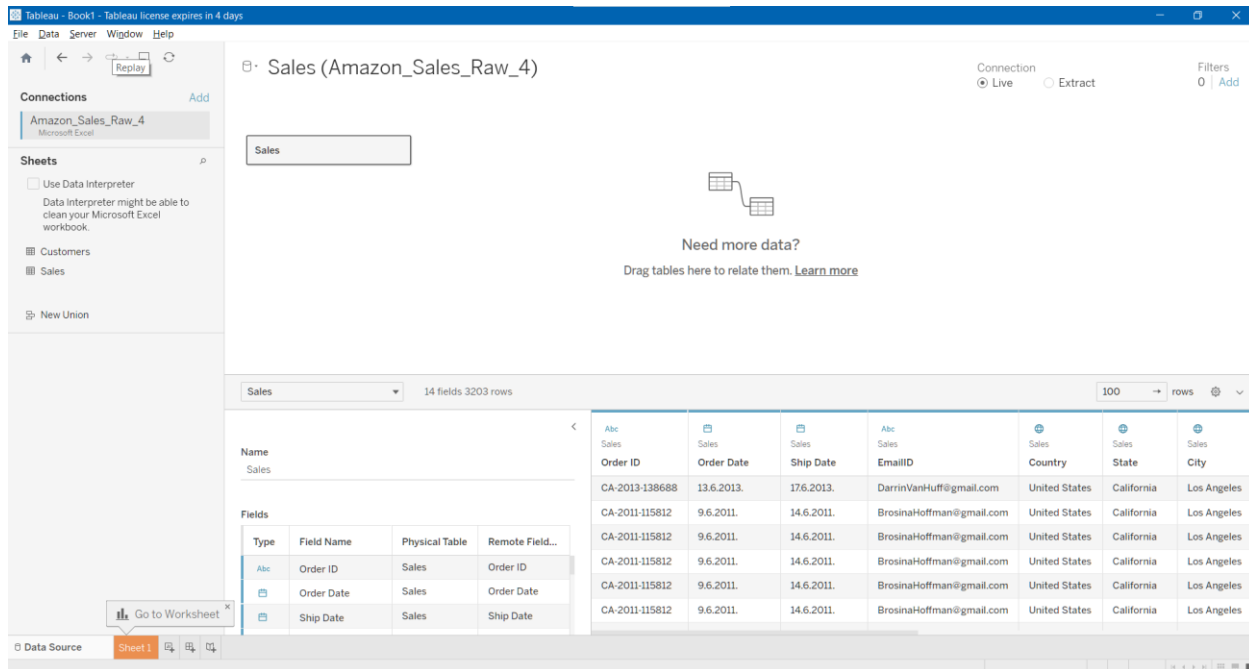
Slika 2 – Amazon Sales dataset nakon uređenja i čišćenja podataka

Proces dodavanja podataka u Tableau softver je relativno jednostavan i počinje dvostrukim klikom na Tablea ikonicu na desktopu, koja se pojavljuje nakon instalacije Tableau programa. Nakon toga, otvara nam se početni ekran gde treba odrediti izvor i format podataka za analizu. Kada izaberemo odgovarajući fajl, otvara nam se prozor gde se u donjem desnom delu ekrana vide kolone dataseta, a levo nam se prikazuju sve tabele dataseta. U ovom prozoru možemo vršiti različite operacije nad tabelama dataseta: promena imena kolona, dodela odgovarajućeg tipa podataka svakoj koloni, sortiranje podataka, određivanje kardinalnosti veze između tabela, JOIN operacije između tabela, i druge. Na početnom ekranu takođe možemo podešavati postojeću konekciju sa izvorom podataka, koristiti Data interpreter program za automatsko čišćenje podataka, filtrirati podatke, itd.

## IV. Spajanje različitih tabela kreiranjem zajedničkog ključa u Tableau softveru

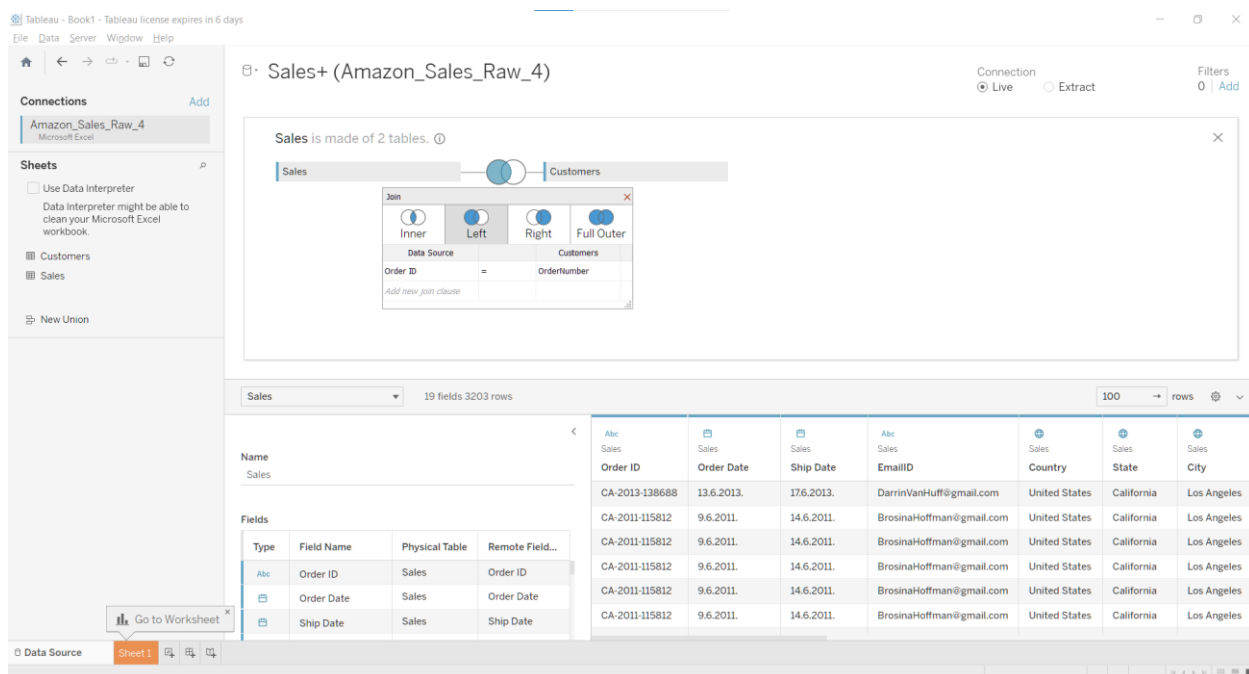
Kada startujemo Tableau i izaberemo izvor podataka za obradu, Tableau na levoj strani ekrana prikazuje tabele izabranog dataseta nad kojima možemo vršiti različite JOIN operacije i definisati kardinalnost veze. „Prevlačenjem“ sa leve na desnu stranu biramo

tabele koje povezujemo kreiranjem zajedničkog ključa. Prvo, drag-and-drop akcijom miša na radnu površinu Tableaua smestamo tabelu *Sales*.



Slika 3 – Početni prozor Tableau programa sa izabranom Sales tabelom

Zatim, dvostrukim levim klikom na tabelu *Sales*, započinjemo spajanje tabela. Levim klikom miša „prevlačimo“ tabelu *Customers*, i pozicioniramo je desno u odnosu na tabelu *Sales*. Tableau prepoznaje proceduru spajanja i automatski pokreće dialog box u kojem definišemo osobine JOIN operacije: vrsta JOIN-a (INNER JOIN, LEFT OUTER JOIN, RIGHT OUTER JOIN, FULL OUTER JOIN) i biranje kolone, odnosno zajedničkog ključa koji povezuje dve tabele. U našem slučaju, odabrali smo LEFT OUTER JOIN operaciju, kako bi obuhvatili sve podatke iz tabele *Sales* i odgovarajuće podatke iz tabele *Customers*. Zajednički ključ pri povezivanju tabela je kolona *Order ID* iz tabele *Sales*, i kolona *OrderNumber* iz tabele *Customers*. Pojavljuje se simbol obojenog levog kruga i delimično obojenog desnog kruga koji označava uspešnu LEFT OUTER JOIN operaciju. U donjem desnom uglu formira se nova prikaz, koji prikazuje sve redove iz tabele *Sales* i one redove iz tabele *Customers* čija se vrednost primarnog zajedničkog ključa poklapa sa vrednošću zajedničkog ključa iz tabele *Customers*.

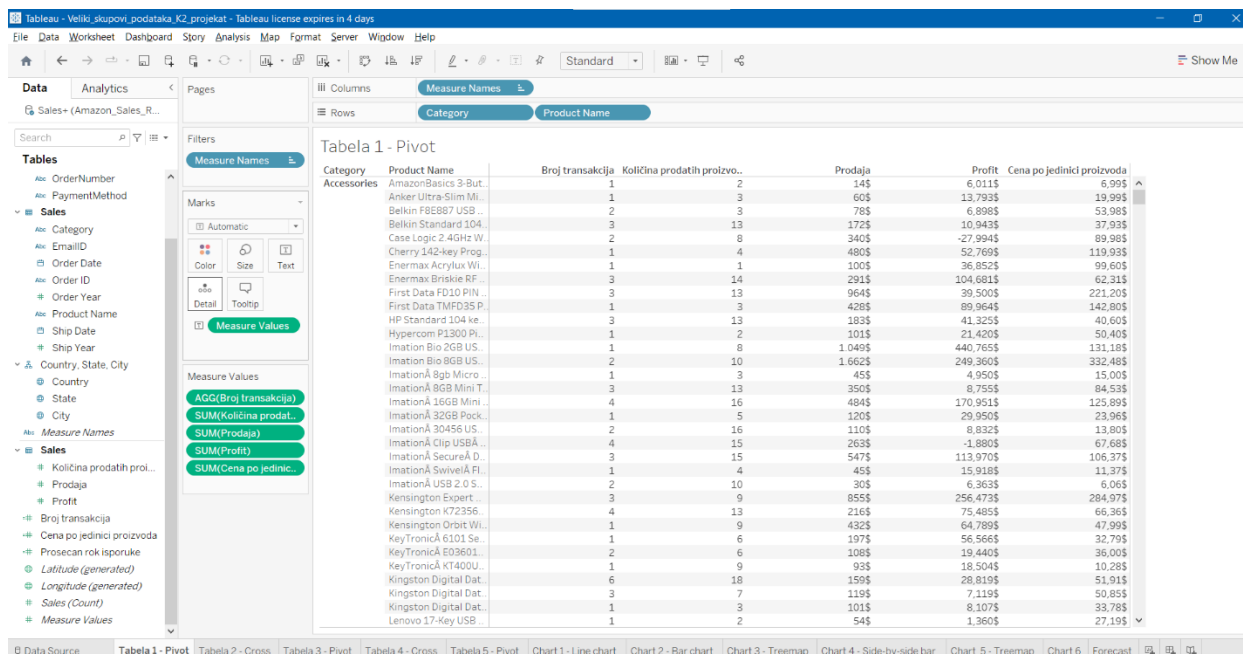


Slika 4 - Uspešno povezivanje tabela Sales i Customers zajedničkim ključem Order ID(OrderNumber)

## V. Kreiranje kros i pivot tabela u Tableau Softveru

Nakon što smo uspešno učitali dataset, povezali tabele korišćenjem zajedničkog ključa, otvara nam se novi prozor namenjen analitici i vizualizaciji podataka. Mali deo ekrana sa leve strane zauzima meni, podeljen na dva dela, *Data i Analytics*. U sekciji *Analytics* nalaze se različite opcije namenjeni analizi podataka: pomoćni alati za sumiranje podataka računanjem konstanti i prosečnih vrednosti, opcija za računanje mediane, alati za modelovanje (Forecast, Cluster, Trend Line), kreiranje referentnih linija, distribucije, itd. U sekciji *Data* prikazane su sve kolone dataseta i njihov tip podataka, podeljene na kategoričke kolone i kolone sa numeričkim vrednostima nad kojima je moguće primenjivati grupne funkcije (AVG, SUM, MIN, MAX, itd.). U posebnoj sekciji nalaze se kalkulacije (numeričke kolone kao što su Sales, Profit) koje olakšavaju analizu, a korisnici mogu dodati i sopstvena računanja klikom na opciju *Create Calculated Field*. Radnu površinu Tableaua uglavnom zauzima centralni pano za vizualnu reprezentaciju podataka. U gornjem delu ekrana, drag-and-drop akcijama možemo definisati kolone i redove „prevlačenem“ u Columns i Rows sekcije, dok se sa leve strane panoa nalaze sekcije Pages, Filters i Marks kao dodatne opcije u procesu vizualizacije. U gornjem desnom uglu, klikom na opciju Show Me, otvaraju nam se različite vrste

grafika(chartova), koje nam Tableau automatski predlaže ili klikom miša možemo sami izabrati odgovarajuću vrstu charta. Tabele i chartovi se kreiraju na isti način, drag-and-drop akcijama miša pozicioniramo kolone i izračunate vrednosti u Columns ili Rows sekciju, a dalja podešavanja vršimo „prevlačenjem“ kolona u Filters ili Marks sekciju. Tableau automatski kreira tabelu ili chart, a u sekciji Show Me možemo odabrati vrstu željenog grafičkog prikaza.

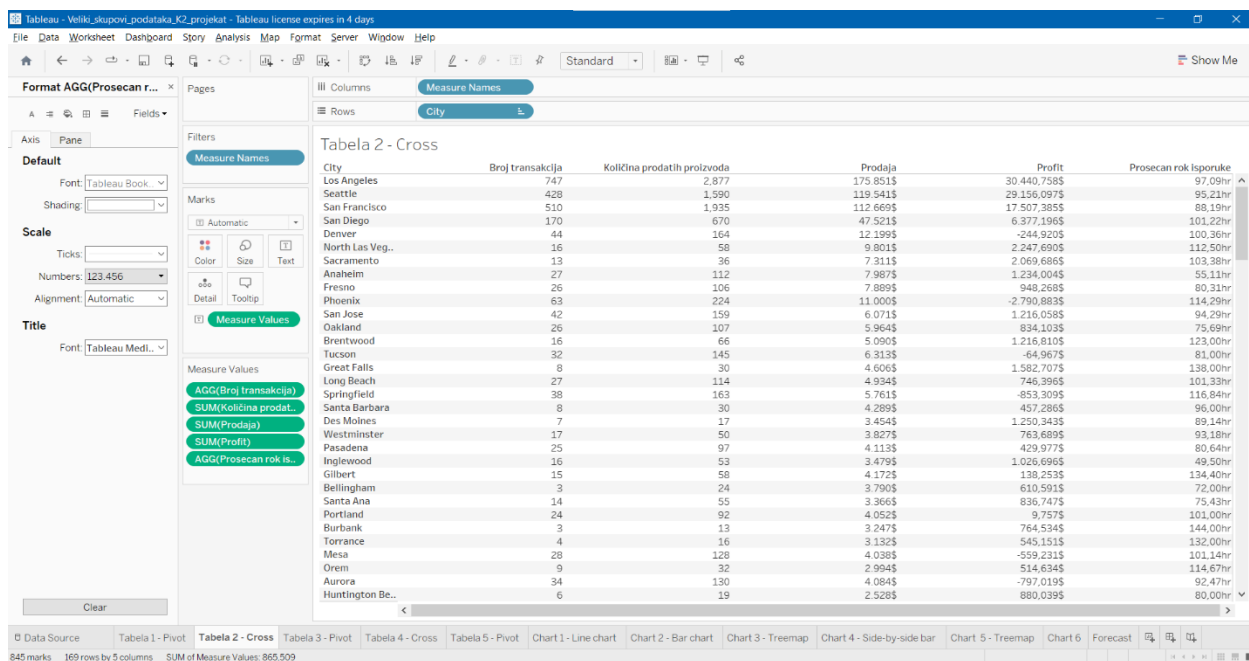


Slika 5 - Radna površina Tableau softverskog alata na kojoj je prikazana Tabela 1

Projekat sadrži pet Cross i Pivot tabela i šest grafika ili chartova. Prva tabela, nazvana Tabela 1, je Pivot tabela koja predstavlja kategorije proizvoda i svaki pojedinačni proizvod, i informacije o uspešnosti prodaje svakog od njih, odnosno broj transakcija u koje je bio uključen proizvod, učinak prodaje, koliki je profit po svakom proizvodu i kolika je cena jedne jedinice proizvoda.

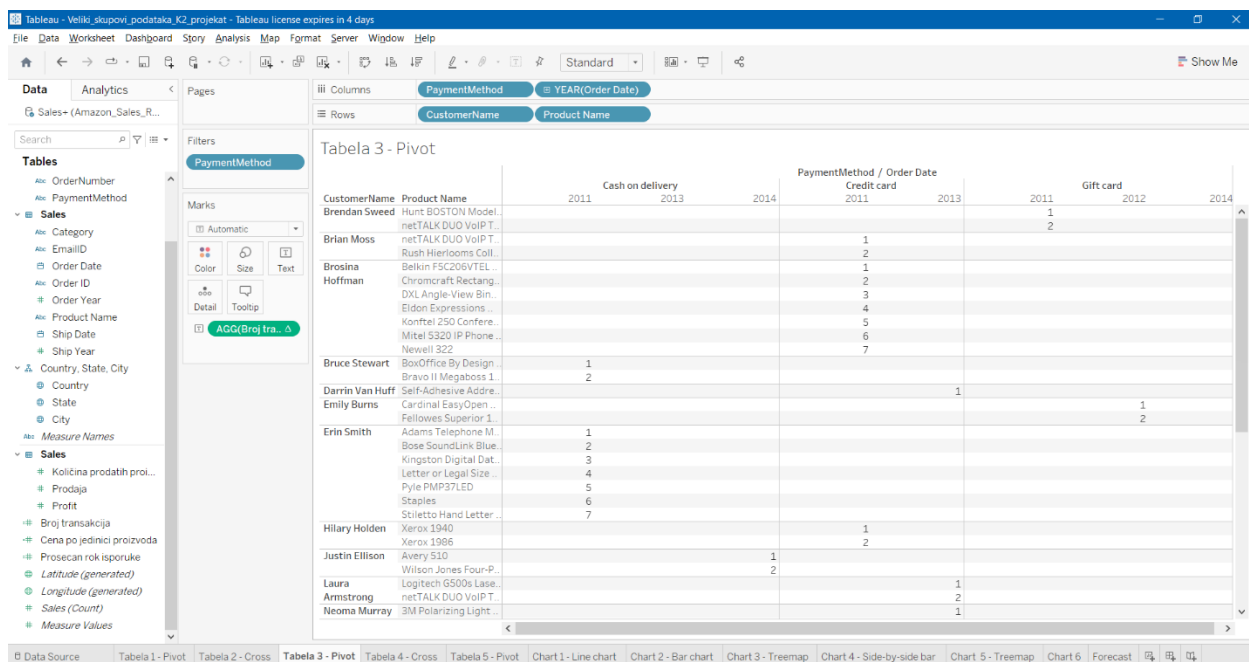
Druge tabela, Tabela 2, je Cross tabela koja predstavlja učinak prodaje prikazan po gradovima, odnosno koliko je transakcija obavljeno iz kog grada, koliko je proizvoda isporučeno i u koji grad, i koji grad je zabeležio najbolje finansijske rezultate prodaje i profita.





Slika 6 – Cross Tabela 2 koja prikazuje statistike o prodaji po gradovima isporuke

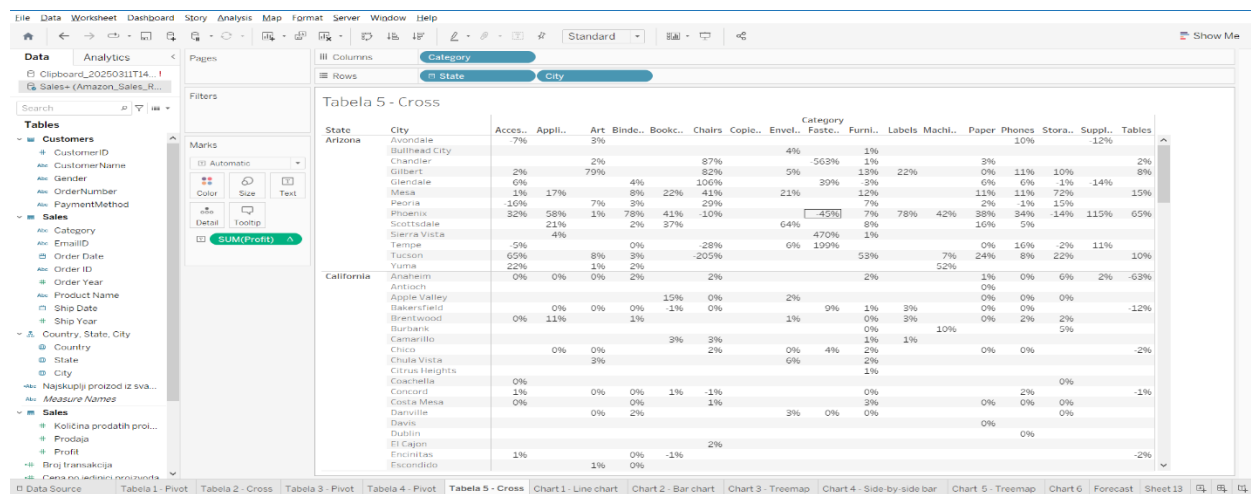
Treća tabela, Tabela 3, je Pivot tabela koja prikazuje imena kupaca, pojedinačne proizvode, godinu transakcije i metod plaćanja koji je korišćen prilikom kupovine: plaćanje kreditnom karticom (Credit card), plaćanje poklon karticom ili kuponom (Gift card) i plaćanje pouzećem (Cash on delivery). Tabela prikazuje sve kupovine jednog kupca i njihov redosled, kao i godinu koje je kupac izvršio transakciju.



Slika 7 – Tabela 3 prikazuje imena kupaca, kupljeni proizvod, način plaćanja i godinu kupovine

[illegible]

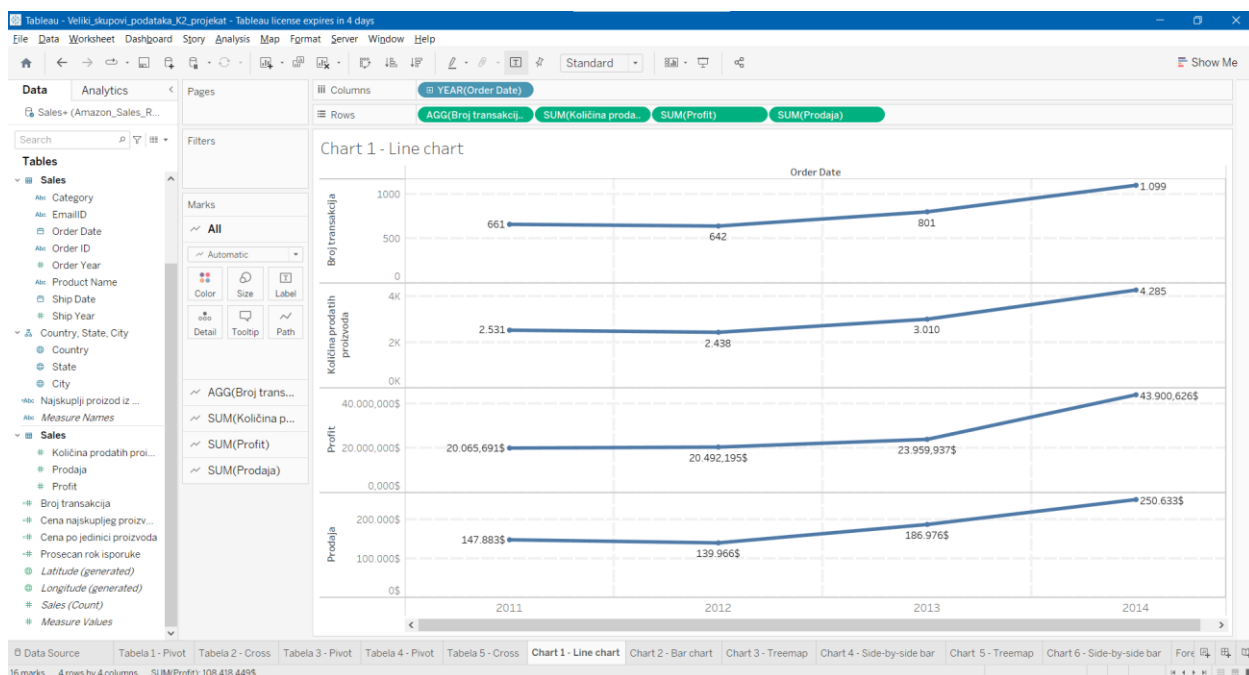
Peta tabela, Tabela 5, je Pivot tabela koja prikazuje koliki je profit kompanija ostvarila za svaku kategoriju proizvoda, a taj je profit je podeljen po lokaciji kupca koji je izvršio kupovinu. Prikazan je procentualni deo profita za svaku kategoriju proizvoda, podeljen po gradovima u kojima je izvršena narudžbina.



Slika 9 – Tabela 5 prikazuje profita ostvaren za svaku kategoriju podeljen po gradovima

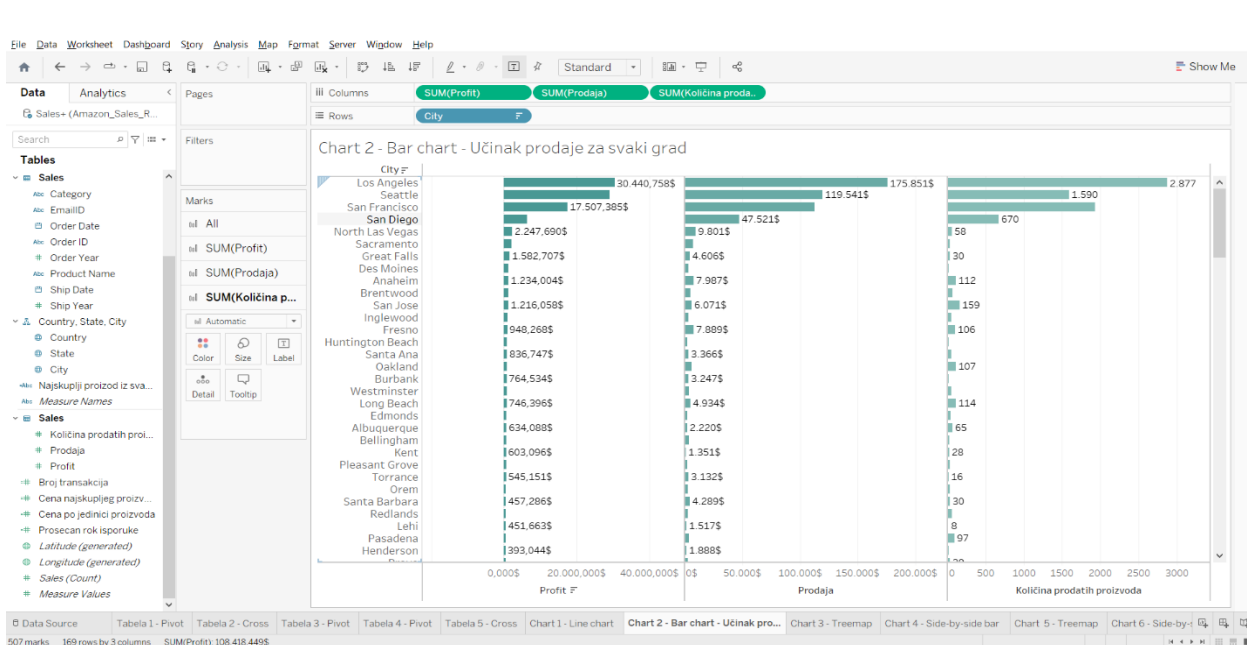
## VI. Kreiranje grafičkih izveštaja u Tableau softveru

Projekat sadrži ukupno 6 grafika ili chartova. Prvi grafik, Chart 1, prikazuje kretanje broja kupovina, količine kupljenih proizvoda, učinka profita i prodaje kroz vreme. Sa grafika primećujemo da su količina kupljenih proizvoda i broj transakcija doživele mali pad u 2012. godinu u odnosu na 2011., dok su naredne dve godine bile godine rasta prodaje, naročito 2014. godina u kojoj je zabeležen najveći rast. Učinak prodaje je u 2014. naglo porastao u odnosu na 2013., dok je profit u 2014. godini dvostruko veći u odnosu na 2013. godinu.



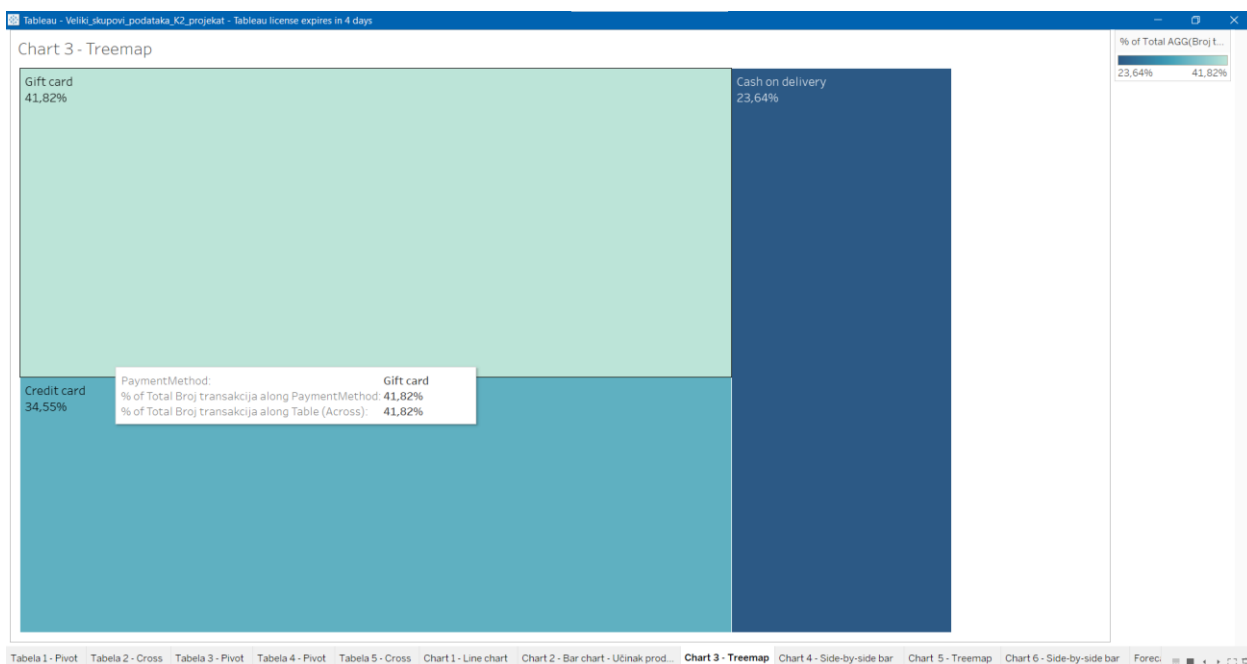
Slika 10 - Chart 1 - Line chart koji predstavlja učinak prodaje, profita, broj transakcija kroz vreme

Na drugom grafiku predstavljene su statistike profita, prodaje i količine prodatih proizvoda za svaki grad iz kojeg je izvršena narudžbina. Ako sortiramo grafik opadajuće, primetićemo da većina prihoda profita dolazi iz tri grada: Los Angeles i San Francisco, najveći gradovi države California, i najveći grad države Washington, Seattle. Sa grafika opažamo da su još prisutni veliki i prosperitetni gradovi kao što su Las Vegas, Phoenix, San Jose, Salt Lake City, Boise, pa bi bilo preporučeno da se strategija privlačenja novih kupaca fokusira na navedene gradove i države.



Slika 11 – Chart 2 - Učinak prodaje prikazan po gradovima

Treći grafik potiče iz kategorije Treemap chartova, i prikazuje načine plaćanja koje su kupci koristili i njihov odnos. Plaćanja su uglavnom vršena koristeći kreditne i poklon kartice, dok je manji deo kupaca koristio mogućnost plaćanja pouzecom.



Slika 12 – Chart 3 prikazuje procentualan odnos načina plaćanja

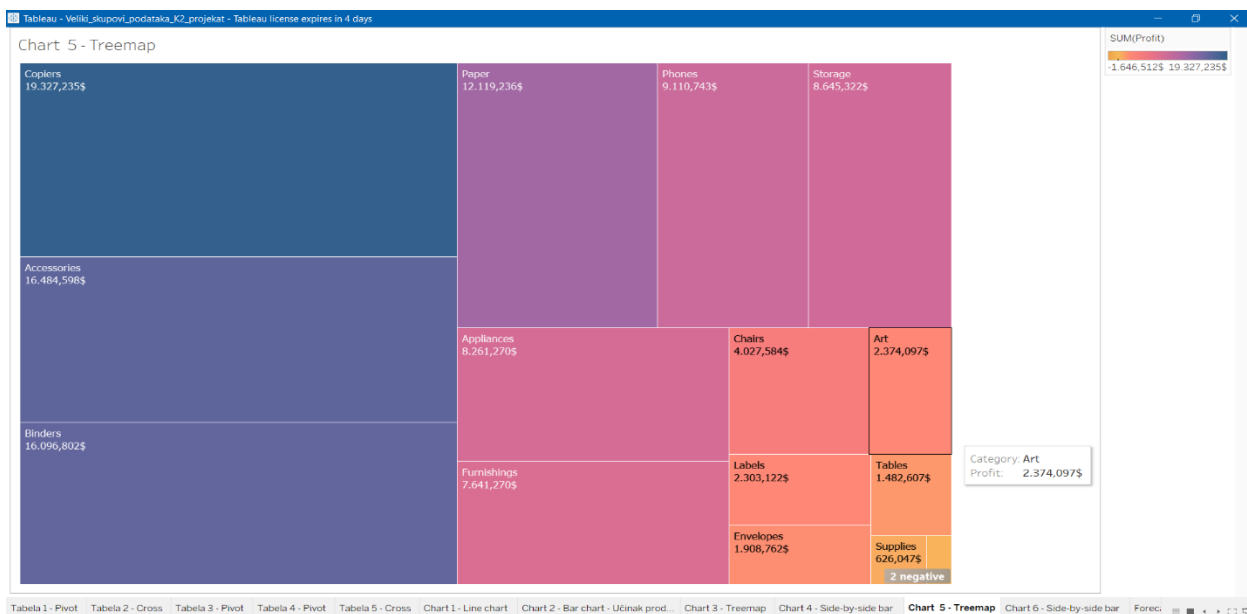
Četvrti grafik, nazvan Chart 4, je vrsta side-by-side grafika i predstavlja prosečan rok isporuke proizvoda za svaku državu isporuke, prikazan u satima. Najveće trajanje

isporuke proizvoda zabeleženo je u državama Wyoming, Montana i New Mexico, a najmanje u državama Idaho, Colorado i California.



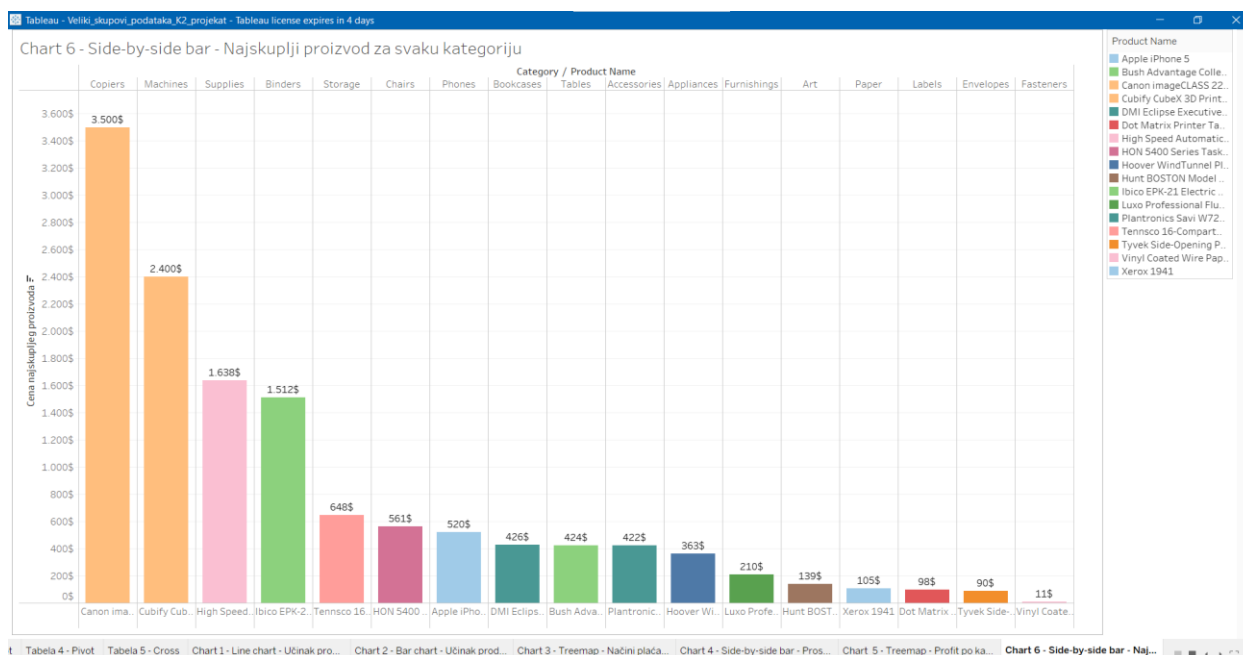
Slika 13 – Chart 4 prikazuje prosečno vreme trajanja isporuke za svaku državu

Peti chart je grafički prikaz iz kategorije Treemap, i prikazuje raspodelu profita po kategorijama proizvoda. Najveći profit doneli su nam proizvodi iz kategorija Copiers , Accessories i Binders, a najmanje proizvodi iz kategorija Fasteners i Supplies.



Slika 14 – Chart 5 prikazuje profit po kategorijama proizvoda

Šesta grafička prezentacija, Chart 6 iz kategorije side-by-side bar grafika prikazuje najskuplji proizvod iz svake kategorije proizvoda, kao i njegovu cenu. Posmatrajući grafik primećujemo da je najskuplji proizvod Canon imageClass 220 Advanced Copier iz kategorije Copiers, koja je istovremeno i kategorija proizvoda u kojoj smo ostvarili najveći profit. Drugi najskuplji proizvod je Cubify CubeX 3D Printer Double Head Printer, iz kategorije Machines. Međutim, kada pogledamo prethodni grafik vidimo da kategorija Machines beleži negativnu vrednost profita, odnosno gubitak od 618,926\$, što zahteva da znatno poboljšamo ponudu proizvoda iz kategorije Machines. Takođe, zapažamo da je treći najskuplji proizvod High Speed Automatic Letter Opener iz kategorije Supplies, koja je u odnosu na druge donela relativno nizak profit od 626,047\$.

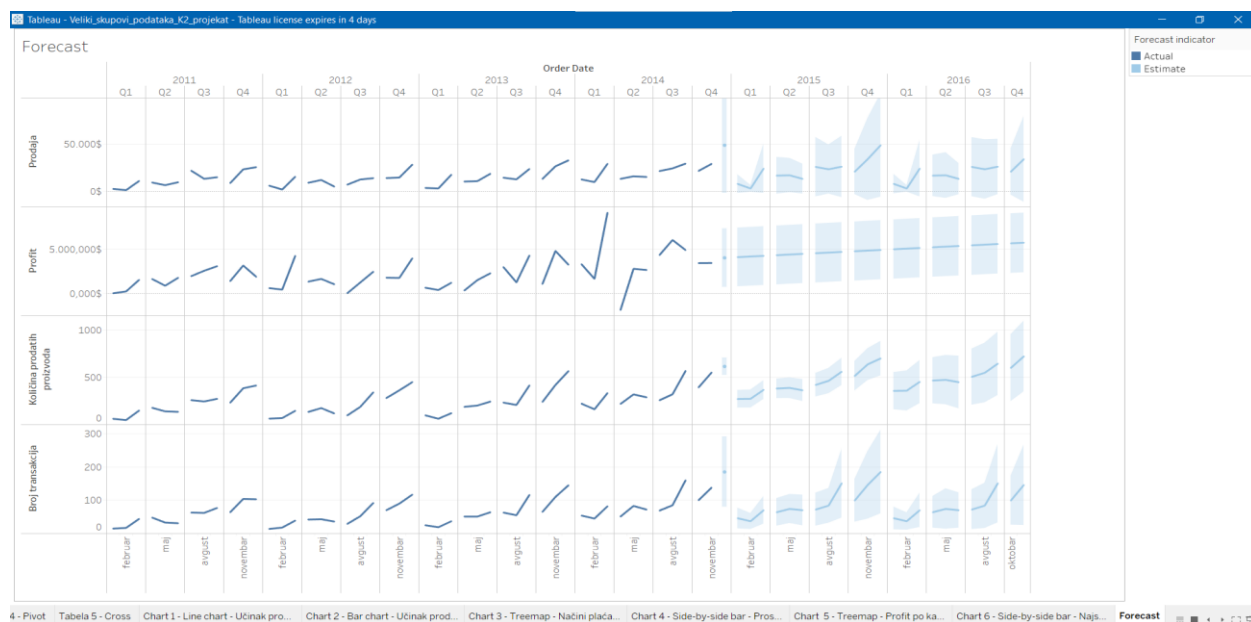


Slika 15 – Najskuplji proizvod iz svake kategorije i njegova cena

## VII. Predviđanje i klasterovanje podataka u Tableau softveru

Alat za predviđanje (eng. Forecasting) u Tableau omogućava nam da na osnovu jedne dimenzije i jedne ili više numeričkih mera vršimo predviđanje vrednosti mera u nekom budućem vremenskom periodu. Da bismo omogućili Forecasting, potrebna nam je kolona sa Date tipom podataka, što u našem projektu predstavlja kolona *Order Date*. „Prevlačimo“ kolonu *Order Date* u Columns sekciju Tableau radne površine. U Rows

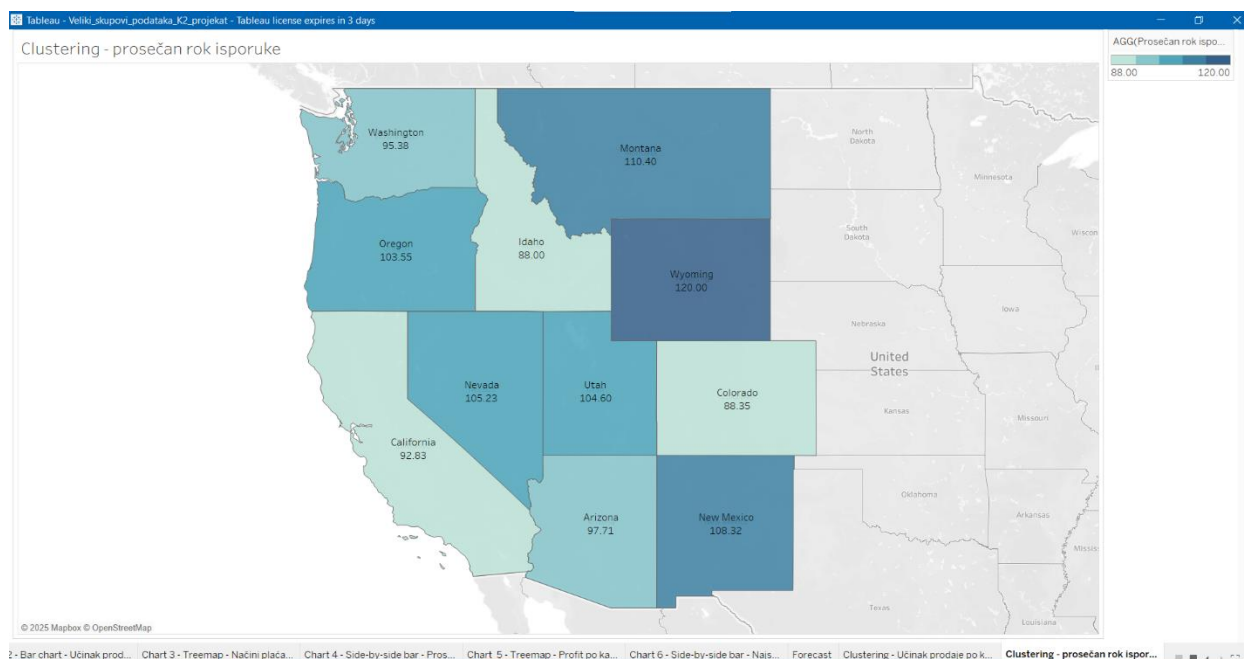
sekciju pozicioniramo dimenzije čiju vrednosti želimo da predviđamo, koje su u slučaju našeg projekta Profit, Prodaja, Količina prodatih proizvoda i Broj transakcija. U Analytics sekciji biramo opciju Forecast, tako što je drag-and-drop akcijom miša pozicioniramo na radnu površinu. Tableau automatski vrši predikciju i izračunava vrednost za budući period od 12 meseci (za potrebe našeg projekta, interval predviđanja podešen je na dve godine).



Slika 16 – Predviđanje(Forecast) različitih vrednosti u Tableau

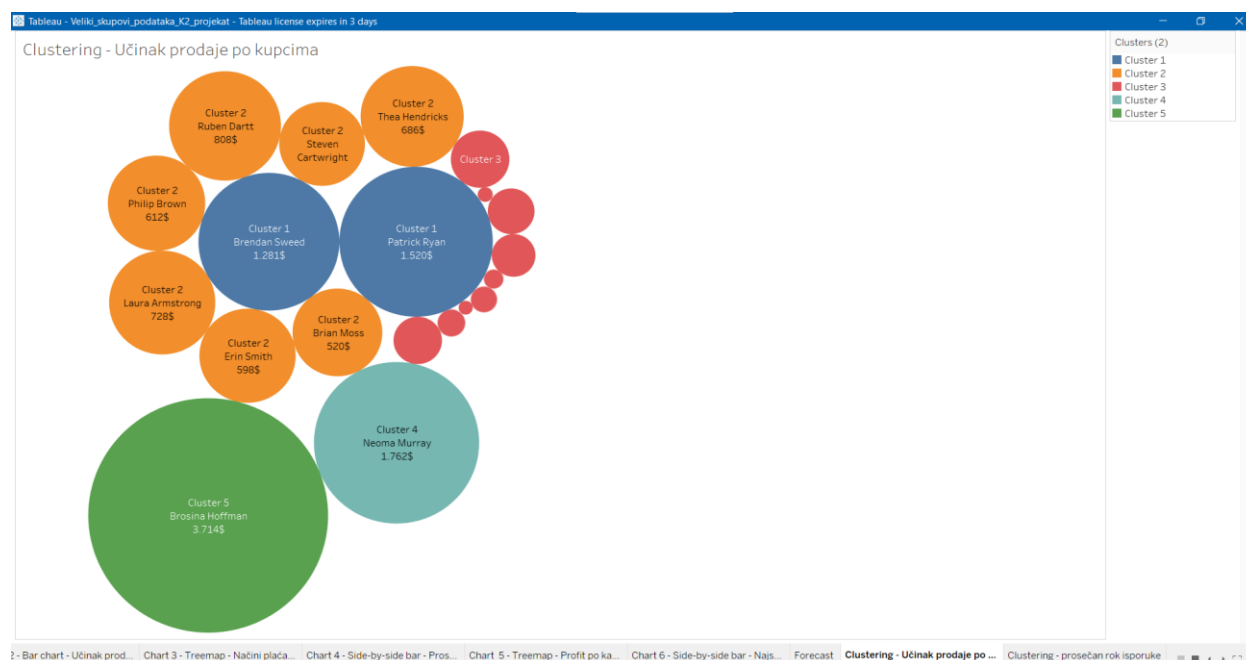
Kreiranje klastera je proces deljenja podataka na male grupe (klaster), gde su podaci unutar klastera grupisani po sličnim i bliskim vrednostima. Proces kreiranja klastera u Tableau softveru je sličan proceduri predviđanja(Forecast), gde se u sekcijama Columns i Rows definišu dimenzije podataka koje želimo da grupišemo. Nakon toga, u sekciji Analytics biramo opciju Cluster i drag-and-drop akcijom miša prevlačimo ikonicu Cluster na radnu površinu. Zatim, otvara se dialog box gde nas Tableau informiše podacima koji će biti deo klasterovanja i nudi opciju da sami definišemo broj klastera ili će softver to učiniti automatski. Rezultat klasterovanja su prikazani podaci grupisani u klaster(grupe) sličnih podataka i obeleženi različitim bojama.

Prvo grupisanje podataka u projektu započeto je pozicioniranjem kolone *State* u Rows sekciju i smeštanjem kalkulacije *Prosečan rok isporuke* u sekciju Columns. Ovim postupkom Tableau prikazuje prosečan rok isporuke za svaku državu u cross tabeli. Otvaramo sekciju *Analytics* i drag-and-drop akcijom miša prevlačimo ikonicu *Cluster* na radnu površinu i tako vršimo klasterovanje podataka.



Slika 17 – Grupisanje podataka o prosečnom roku isporuke proizvoda za svaku državu prikazano na karti

Drugi postupak klasterovanja u projektu je prikazivanje učinka prodaje po svakom kupcu. Procedura klasterovanja je slična prethodnoj, pozicioniramo odgovarajuće kolone (CustomerName i Prodaja) u Columns i Rows sekcije. Ikonicu Cluster drag-and-drop akcijom miša prevlačimo na radnu površinu što navodi Tableau da izvrši grupisanje podataka.



Slika 17 – Podaci o prodaji za svakog kupca grupisani u klastere prikazani Circles chartom



U prvom postupku grupisanja podataka nastalo je 4 klastera država, a najveći broj članova imaju klasteri 1 i 2. Klaster 4 sadrži samo jednog člana (državu Wyoming), a klaster 3 dva (države New Mexico i Montana). U drugom postupku grupisanja pojavljuje se pet klastera kupaca, gde najviše članova imaju klasteri 2 i 3. Klasteri 5 i 4 imaju po jednog člana (kupce Brosinu Hoffman i Neomu Murray) a klaster 1 dva kupca (Brendana Sweeda i Patricka Ryana).