

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 prenestili 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 Tableaua 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 identificuje 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.

	B	C	D	E	F	G	H	I	J
1	Order Date	Ship Date	EmailID	Geography	Category	Product Name	Sales	Quantity	Profit
2	13.6.2013	17.6.2013	DarinVanHuff@gmail.com	United States,Los Angeles,California	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14,62	2,00	6,67
3	9.6.2011	14.6.2011	BrosinHoffmann@gmail.com	United States,Los Angeles,California	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48,86	7,00	14,17
4	9.6.2011	14.6.2011	BrosinHoffmann@gmail.com	United States,Los Angeles,California	Art	Newell 322	7,28	4,00	1,97
5	9.6.2011	14.6.2011	BrosinHoffmann@gmail.com	United States,Los Angeles,California	Phones	Mitel 5320 IP Phone VoIP phone	907,15	4,00	90,72
6	9.6.2011	14.6.2011	BrosinHoffmann@gmail.com	United States,Los Angeles,California	Binders	DXL Angle-View Binders with Locking Rings by Samsill	18,50	3,00	5,78
7	9.6.2011	14.6.2011	BrosinHoffmann@gmail.com	United States,Los Angeles,California	Appliances	Belkin F5C206VTEL 6 Outlet Surge	114,90	5,00	34,47
8	9.6.2011	14.6.2011	BrosinHoffmann@gmail.com	United States,Los Angeles,California	Tables	Chromcraft Rectangular Conference Tables	1706,18	9,00	85,31
9	9.6.2011	14.6.2011	BrosinHoffmann@gmail.com	United States,Los Angeles,California	Phones	Kontrol 250 ConferenceA phoneA - Charcoal black	911,42	4,00	68,36
10	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
11	13.5.2011	15.5.2011	AlejandroGrove@gmail.com	United States,West Jordan,Utah	Storage	Fellowes Super Stor/Drawer	55,50	2,00	9,99
12	27.8.2011	1.9.2011	ZuschussDonatelli@gmail.com	United States,San Francisco,California	Art	Newell 341	8,56	2,00	2,48
13	27.8.2011	1.9.2011	ZuschussDonatelli@gmail.com	United States,San Francisco,California	Phones	Cisco SPA 501G IP Phone	213,48	3,00	16,01
14	27.8.2011	1.9.2011	ZuschussDonatelli@gmail.com	United States,San Francisco,California	Binders	Wilson Jones Hanging View Binder, White, 1"	22,72	4,00	7,38
15	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	16.1.2013	20.1.2013	EricHoffmann@gmail.com	United States,Los Angeles,California	Binders	Wilson Jones Active Use Binders	11,65	2,00	4,22
17	16.1.2013	20.1.2013	EricHoffmann@gmail.com	United States,Los Angeles,California	Accessories	ImationA 8GB Mini TravelDrive USB 2.0A Flash Drive	90,57	3,00	11,77
18	17.7.2013	23.7.2013	RubenAusmann@gmail.com	United States,Los Angeles,California	Storage	Eldon Base for stackable storage shelf, platinum	77,88	2,00	3,89
19	24.11.2012	30.11.2012	KunstMiller@gmail.com	United States,Los Angeles,California	Accessories	Vermabit 25 GB 8x Blu-ray Single Layer Recordable Disc, 3/Pack	13,98	2,00	6,15
20	24.11.2012	30.11.2012	KunstMiller@gmail.com	United States,Los Angeles,California	Binders	ACO PRESTEX Data Binder with Storage Hooks, Dark Blue, 14 7/8" X 11"	25,82	6,00	9,36
21	24.11.2012	30.11.2012	KunstMiller@gmail.com	United States,Los Angeles,California	Paper	Xerox 1943	146,73	3,00	68,96
22	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
23	5.12.2011	10.12.2011	BrendanSweet@gmail.com	United States,Gilbert,Arizona	Art	Hunt BOSTON Model 1605 High-Volume Electric Pencil Sharpener, Beige	1113,02	8,00	111,30
24	5.12.2011	10.12.2011	BrendanSweet@gmail.com	United States,Gilbert,Arizona	Phones	netTALK DUO VoIP Telephone Service	167,97	4,00	62,99
25	12.10.2011	16.10.2011	DuaneNoonan@gmail.com	United States,San Francisco,California	Art	Premium Writing Pencils, Soft, #2 by Central Association for the Blind	14,90	5,00	4,17
26	12.10.2011	16.10.2011	DuaneNoonan@gmail.com	United States,San Francisco,California	Storage	Sortfile Multipurpose Personal File Organizer, Black	21,39	1,00	6,20
27	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
28	18.9.2013	23.9.2013	JimSink@gmail.com	United States,Los Angeles,California	Phones	Panasonic Kx-TS550	73,58	2,00	8,28
29	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.

EmailID	Country	State	City	Category	Product Name	Sales	Quantity	Profit	Order Year	Ship Date
DarvinVanHuff@gmail.com	United States	California	Los Angeles	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14,62	2	6,8714	2013	
BrosinaHoffman@gmail.com	United States	California	Los Angeles	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48,86	7	14,1694	2011	
BrosinaHoffman@gmail.com	United States	California	Los Angeles	Art	Newell 322	7,28	4	1,9656	2011	
BrosinaHoffman@gmail.com	United States	California	Los Angeles	Phones	Mitel 5320 IP Phone VoIP phone	907,152	4	90,7152	2011	
BrosinaHoffman@gmail.com	United States	California	Los Angeles	Binders	DXL Angle-View Binders with Locking Rings by Samsill	18,504	3	5,7625	2011	
BrosinaHoffman@gmail.com	United States	California	Los Angeles	Appliances	Belkin FSG206VTEL 6 Outlet Surge	114,9	5	34,47	2011	
BrosinaHoffman@gmail.com	United States	California	Los Angeles	Tables	Chromcraft Rectangular Conference Tables	1706,184	9	85,3092	2011	
BrosinaHoffman@gmail.com	United States	California	Los Angeles	Phone	Konfet 250 Conference phone - Charcoal black	911,424	4	68,3568	2011	
IreneMaddon@gmail.com	United States	Washington	Seattle	Binders	Fellowes PB200 Plastic Comb Binding Machine	407,976	3	132,5922	2013	
AlejandroGrove@gmail.com	United States	Utah	West Jordan	Storage	Fellowes Super Stor/Drawer	55,5	2	9,99	2011	
ZuschussDonatelli@gmail.com	United States	California	San Francisco	Art	Newell 341	8,56	2	2,4824	2011	
ZuschussDonatelli@gmail.com	United States	California	San Francisco	Phones	Cisco SPA 501G IP Phone	213,48	3	16,011	2011	
ZuschussDonatelli@gmail.com	United States	California	San Francisco	Binders	Wilson Jones Hanging View Binder, White, 1"	22,72	4	7,384	2011	
EmilyBurns@gmail.com	United States	Utah	Orem	Tables	Bretford CR4500 Series Slim Rectangular Table	1044,63	3	240,2649	2012	
EricHoffmann@gmail.com	United States	California	Los Angeles	Binders	Wilson Jones Active Use Binders	11,648	2	4,2224	2013	
EricHoffmann@gmail.com	United States	California	Los Angeles	Accessories	Imation 4 GB Mini TravelDrive USB 2.0 Å Flash Drive	90,57	3	11,7741	2013	
RubenAusman@gmail.com	United States	California	Los Angeles	Storage	Eldon Base for stackable storage shelf, platinum	77,88	2	3,894	2013	
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	
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	
KunstMiller@gmail.com	United States	California	Los Angeles	Paper	Xerox 1943	146,73	3	68,9631	2012	
KunstMiller@gmail.com	United States	California	Los Angeles	Furnishings	Luxo Economy Swing Arm Lamp	79,76	4	22,3328	2012	
BrendanSweet@gmail.com	United States	Arizona	Gilbert	Art	Hunt BOSTON Model 1606 High-Volume Electric Pencil Sharpener, Beige	1113,024	8	111,3024	2011	
BrendanSweet@gmail.com	United States	Arizona	Gilbert	Phones	netTALK DUO VoIP Telephone Service	167,968	4	62,988	2011	
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	
DuaneNoonan@gmail.com	United States	California	San Francisco	Storage	Sortfile Multipurpose Personal File Organizer, Black	21,39	1	6,2031	2011	
JimSink@gmail.com	United States	California	Los Angeles	Art	Turquoise Lead Holder with Pocket Clip	20,1	3	6,633	2013	
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: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 tabeli 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 smeštamo tabelu *Sales*.

Order ID	Order Date	Sales	Ship Date	EmailID	Sales Country	Sales State	Sales City
CA-2013-138688	13.6.2013.		17.6.2013.	DarrinVanHuff@gmail.com	United States	California	Los Angeles
CA-2011-115812	9.6.2011.		14.6.2011.	BrosinaHoffman@gmail.com	United States	California	Los Angeles
CA-2011-115812	9.6.2011.		14.6.2011.	BrosinaHoffman@gmail.com	United States	California	Los Angeles
CA-2011-115812	9.6.2011.		14.6.2011.	BrosinaHoffman@gmail.com	United States	California	Los Angeles
CA-2011-115812	9.6.2011.		14.6.2011.	BrosinaHoffman@gmail.com	United States	California	Los Angeles

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*.

The screenshot shows the Tableau interface with the following details:

- Top Bar:** Tableau - Book1 - Tableau license expires in 6 days, File, Data, Server, Window, Help.
- Left Sidebar (Connections):** Amazon\_Sales\_Raw\_4 (Microsoft Excel), Sheets: Customers, Sales, New Union, Use Data Interpreter (unchecked).
- Central Area:**
  - Sales+ (Amazon\_Sales\_Raw\_4) Sheet:** Sales is made of 2 tables. A join dialog shows Sales (Order ID) joined with Customers (OrderNumber) using an Inner Join.
  - Data View:** Sales sheet, 19 fields, 3203 rows. Fields include Order ID, Order Date, Ship Date, Sales, EmailID, Country, State, City.
- Bottom:** Data Source, Sheet1, various icons.

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 sumiranju 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 Tableaura uglavnom zauzima centralni pano za vizuelnu 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 izabrat odgovarajuću vrstu charta.Tabele i chartovi se kreiraju na isti načina, drag-and-drop akcijama miša pozicioniramo kolone i izračunate vrednosti u Columns ili Rows sekiju, a dalja podešavanja vršimo „prevlačenjem“ kolona u Filters ili Marks sekiju.Tableau automatski kreira tabelu ili chart, a u sekiji Show Me možemo odabrati vrstu željenog grafičkog prikaza.

Category	Product Name	Broj transakcija	Količina prodatih proizvoda	Prodaja	Profit	Cena po jedinici proizvoda
Accessories	AmazonBasics 3-But...	1	2	14\$	6,011\$	6,99\$
	Anker Ultra-Slim Mi...	1	3	60\$	13,793\$	19,99\$
	Belkin FEBB87 USB...	2	3	78\$	6,898\$	53,98\$
	Belkin Standard 104...	3	13	172\$	10,943\$	37,93\$
	Case Logic 2.4GHz W...	2	8	340\$	-27,994\$	89,30\$
	Cherry 142-Key Prog...	1	4	480\$	52,769\$	119,93\$
	Enermax Acrylux Wi...	1	1	100\$	36,852\$	99,60\$
	Enermax Briskie RF...	3	14	291\$	104,681\$	62,31\$
	First Data FD10 PIN...	3	13	964\$	39,50\$	221,20\$
	First Data TMFD35 P...	1	3	426\$	89,964\$	142,80\$
	HP Stream 14-d000na...	3	13	183\$	11,925\$	40,60\$
	Hyperion F1300 Pi...	1	2	101\$	21,420\$	50,40\$
	Imation Bio 2GB US...	1	8	1,040\$	440,765\$	131,18\$
	Imation Bio 8GB US...	2	10	1,662\$	249,369\$	332,46\$
	Imation 1gb Micro...	1	3	45\$	4,050\$	15,00\$
	Imation 8GB Mini T...	3	13	350\$	8,755\$	88,53\$
	Imation 16GB Mini...	4	16	494\$	170,951\$	125,59\$
	Imation 32GB Peck...	1	5	120\$	29,950\$	23,96\$
	Imation 30456 US...	2	16	110\$	8,823\$	13,80\$
	Imation Clip USB A...	4	15	263\$	-1,880\$	67,68\$
	Imation Secure A D...	3	15	547\$	113,970\$	106,37\$
	Imation Swivel A Fl...	1	4	45\$	15,918\$	11,37\$
	Imation USB 2.0 S...	2	10	30\$	6,363\$	6,06\$
	Kensington Expert...	3	9	855\$	256,473\$	284,97\$
	Kensington K72356...	4	13	216\$	75,485\$	66,36\$
	Kensington Orbit Wi...	1	9	432\$	64,789\$	47,99\$
	KeyTronik E101 Se...	1	6	197\$	56,566\$	32,79\$
	KeyTronik E03601...	2	6	108\$	19,440\$	36,00\$
	KeyTronik KT400U...	1	9	93\$	18,504\$	10,28\$
	Kingston Digital Dat...	6	18	159\$	28,819\$	51,91\$
	Kingston Digital Dat...	3	7	119\$	7,11\$	50,85\$
	Kingston Digital Dat...	1	3	101\$	8,107\$	33,78\$
	Lenovo 17-Key USB...	1	2	54\$	1,360\$	27,19\$

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.

Druga 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.

**Tabela 2 - Cross**

City	Broj transakcija	Količina prodatih proizvoda	Prodaja	Profit	Prosecan rok isporuke
Los Angeles	747	2.877	175.851\$	30.440.758\$	97,09hr
Seattle	428	1.590	119.541\$	29.156.097\$	95,21hr
San Francisco	510	1.935	112.669\$	17.507.385\$	88,19hr
San Diego	170	670	47.521\$	6.377.196\$	101,22hr
Denver	44	164	12.199\$	-24.9.20\$	100,36hr
North Las Veg..	16	58	9.801\$	2.247.690\$	112,50hr
Sacramento	13	36	7.311\$	2.069.686\$	103,38hr
Anaheim	27	112	7.987\$	1.234.004\$	55,11hr
Fresno	26	106	7.889\$	948.268\$	80,31hr
Phoenix	63	224	5.090\$	-2.790.882\$	114,29hr
San Jose	42	159	6.071\$	1.211.553\$	94,29hr
Oakland	26	107	5.955\$	894.103\$	75,58hr
Brentwood	16	66	4.289\$	1.216.610\$	123,00hr
Tucson	32	145	6.513\$	-64.967\$	81,00hr
Great Falls	8	30	4.606\$	1.582.707\$	138,00hr
Long Beach	27	114	4.924\$	746.396\$	101,32hr
Springfield	38	163	5.761\$	-853.309\$	116,64hr
Santa Barbara	8	30	4.289\$	457.286\$	96,00hr
Des Moines	7	17	3.454\$	1.250.343\$	89,14hr
Westminster	17	50	3.827\$	763.689\$	93,18hr
Pasadena	25	97	4.113\$	429.977\$	80,64hr
Inglewood	16	53	3.479\$	1.026.696\$	49,50hr
Gilbert	15	58	4.172\$	138.253\$	134,40hr
Bellingham	3	24	3.790\$	610.591\$	72,00hr
Santa Ana	14	55	3.366\$	836.747\$	75,43hr
Portland	24	92	4.052\$	9.757\$	101,00hr
Burbank	3	13	3.247\$	764.534\$	144,00hr
Torrance	4	16	3.132\$	545.151\$	132,00hr
Mesa	28	128	4.038\$	-559.231\$	101,14hr
Orem	9	32	2.994\$	514.634\$	114,67hr
Aurora	34	130	4.084\$	-797.019\$	92,47hr
Huntington Be..	6	19	2.528\$	880.039\$	80,00hr

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.

**Tabela 3 - Pivot**

CustomerName	Product Name	Cash on delivery			PaymentMethod / Order Date			
		2011	2013	2014	2011	2013	2011	2012
Brendan Swed	Hunt BOSTON Model..				Credit card			
	netTALK DUO VoIP T..				2011		1	
Brian Moss	netTALK DUO VoIP T..				2013		2	
	Rush Hierlooms Coll..							
Brosina	Belkin F5C206VTEL..				1			
Hoffman	Chromcraft Rectang..				2			
	DXL Angle-View Bin..				3			
	Eldon Expressions ..				4			
	Kontrol 250 Confere..				5			
	Mitel 5320 IP Phone..				6			
	Newell 3200				7			
Bruce Stewart	BoxOffice By Design..	1						
	JavaOne Metaphor 1..	2						
Darrin Van Huff	Self-Adhesive Addre..					1		
Emily Burns	Cardinal EasyOpen ..						1	
	Fellowes Superior 1..						2	
Erin Smith	Adams Telephone M..	1						
	Bose SoundLink Blue..	2						
	Kingston Digital Dat..	3						
	Letter or Legal Size..	4						
	Pyle PMP37LED	5						
	Staples	6						
	Stiletto Hand Letter..	7						
Hillary Holden	Xerox 1940				1			
	Xerox 1986				2			
Justin Ellison	Avery 510				1			
	Wilson Jones Four-P..				2			
Laura Armstrong	Logitech G500s Lase..					1		
	netTALK DUO VoIP T..					2		
Neoma Murray	3M Polarizing Light..						1	

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

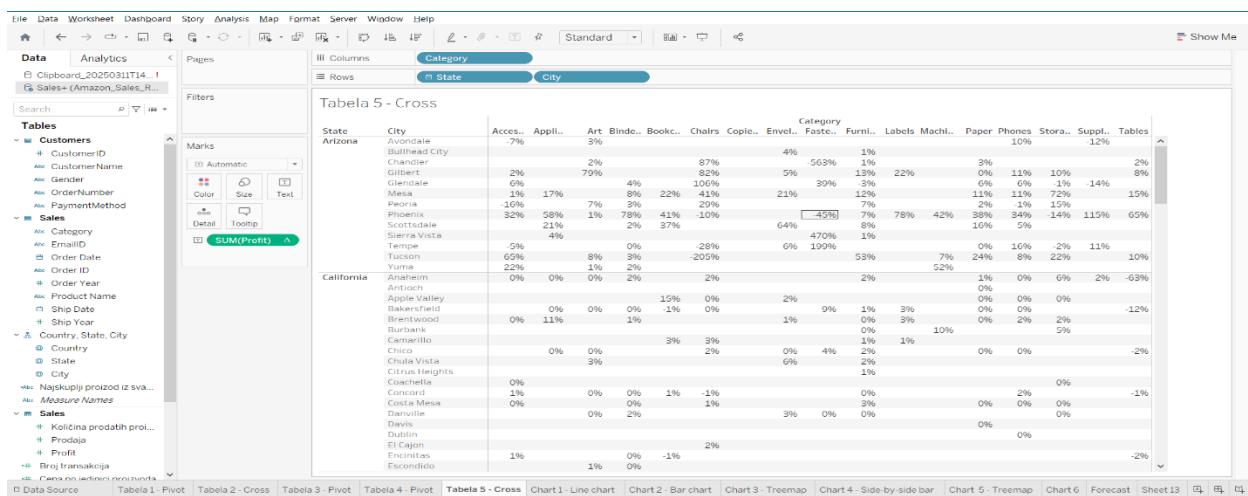
Četvrta tabela prikazuje količinu kupljenih proizvoda podeljenih prema polu, odnosno tačan datum kupovine određene količina nekog proizvoda, i kojoj kategoriji taj proizvod pripada. Kalkulacija tabele sabira količinu kupljenih proizvoda i deli ih prema polu, na kraju prikazujući ukupnu količinu kupljenih proizvoda koju je svaki pol kupio.

**Tabela 4 - Cross**

Gender	Category	8.2.2011.	23.9.2011.	27.9.2011.	2.10.2011.	10.11.2011.	24.11.2011.	5.12.2011.	13.1.2012.	27.4.2012.	23.5.2012.	9.1.2013.	13.6.2013.	11.7.2013.	8.11.2013.	4.12.2013.	2.1.2014.	22.8.2014.
Female	Appliances									16	20	6					20	
	Art									24							24	
	Binders										27						34	
	Bookcases	34																
	Chairs		39															
	Envelopes			45														
	Fasteners			47														
	Furnishings				54	54			54								60	60
	Labels					60							60				60	
	Paper	64	66	68	68											70		
	Phones		78	83	83				83								92	
	Supplies			94							94							103
	Tables																	
Male	Accessories						1									1		
	Appliances															10		
	Art										18					22		
	Binders		47									25					40	
	Bookcases			47														
	Chairs			47														
	Envelopes				47													
	Fasteners				47													
	Furnishings					54	58		71							71	82	86
	Labels					94										96		101
	Paper	102	102	102	110				124				130			110		128
	Phones		112	112	120								124					135
	Supplies			128														
	Tables																	

Slika 8 – Tabela 4 prikazuje količinu kupljenih proizvoda za svaki pol

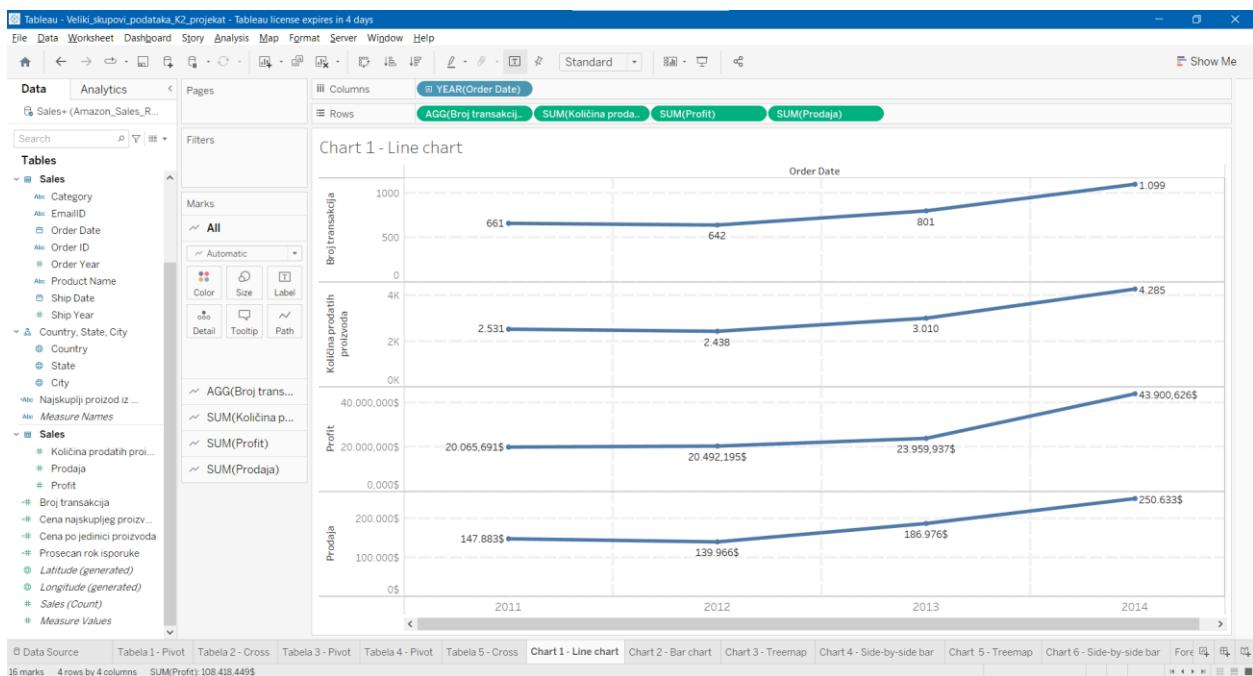
Peta tabela, Tabela 5, je Pivot tabela koja prikazuje koliki je profit kompanija ostvarila za svaku kategoriju proizvoda, a taj je profit je podeljen po lokaciji kupca koji je izvršio kupovinu. Prikazan je prrocentualnu deo profita za svaku kategoriju proizvoda, podeljen po gradovima u kojima je izvršena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



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

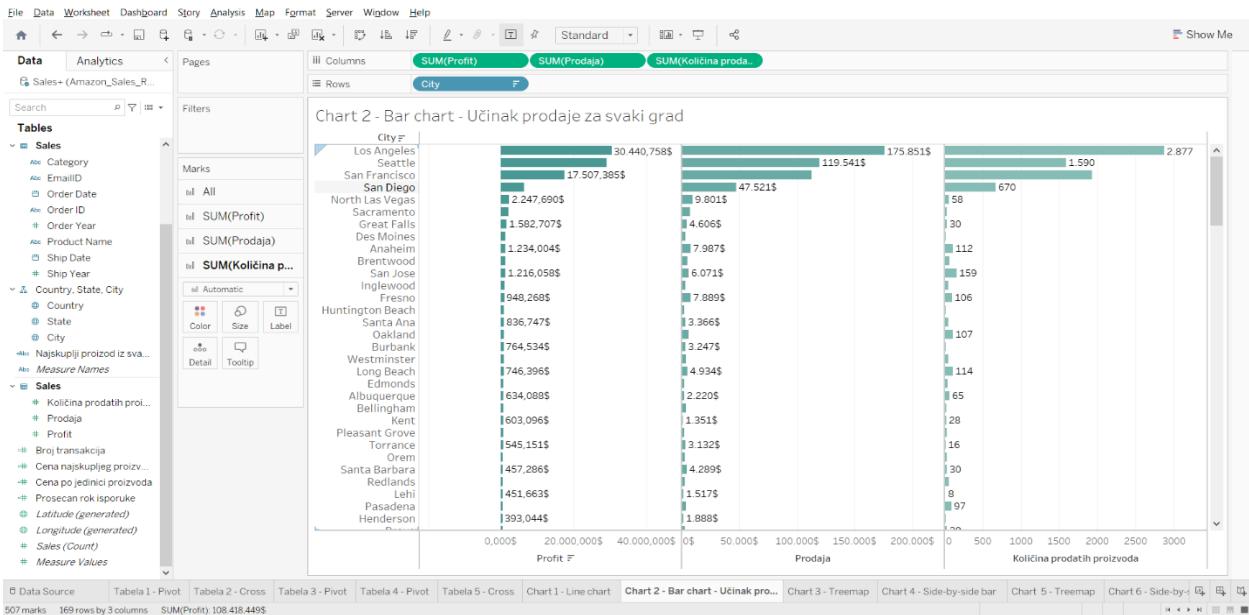
## VI. Kreiranje grafičkih izveštaja u Tableau softveru

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.



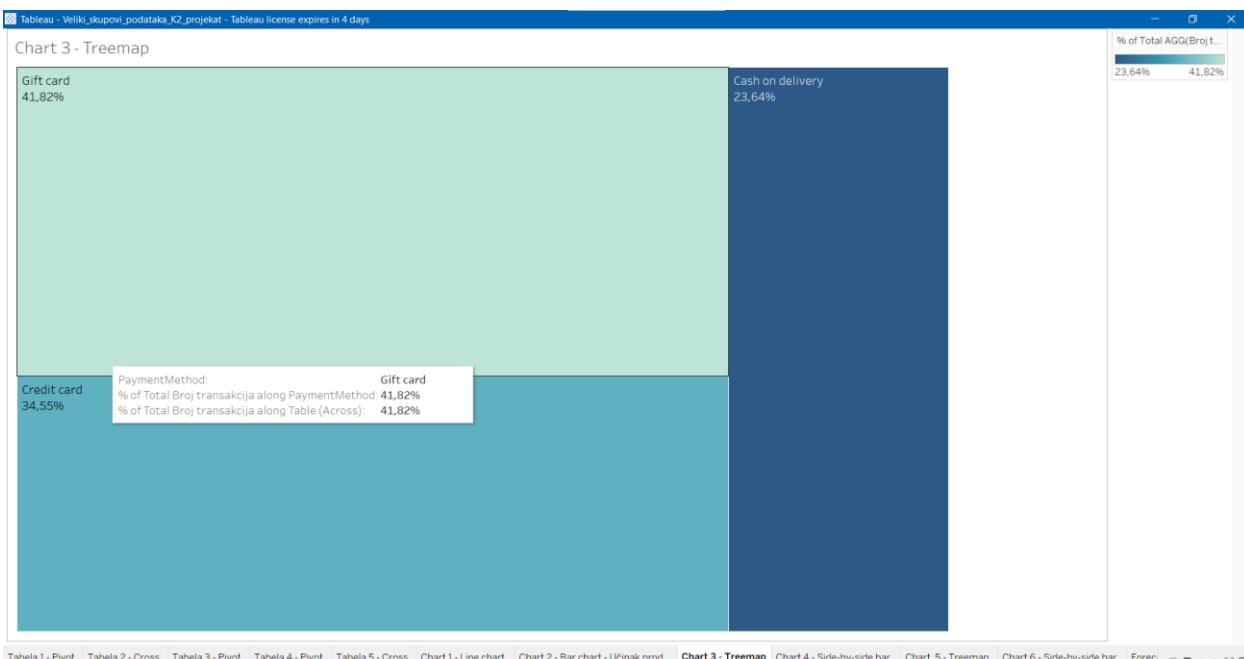
Slika 10 - Chart 1 - Line chart koji predstavlja učinak prodaje, profita, broj transakcija kroz vreme

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 pouzećem.



Slika 12 – Chart 3 prikazuje procentualan odnos načina plaćanja

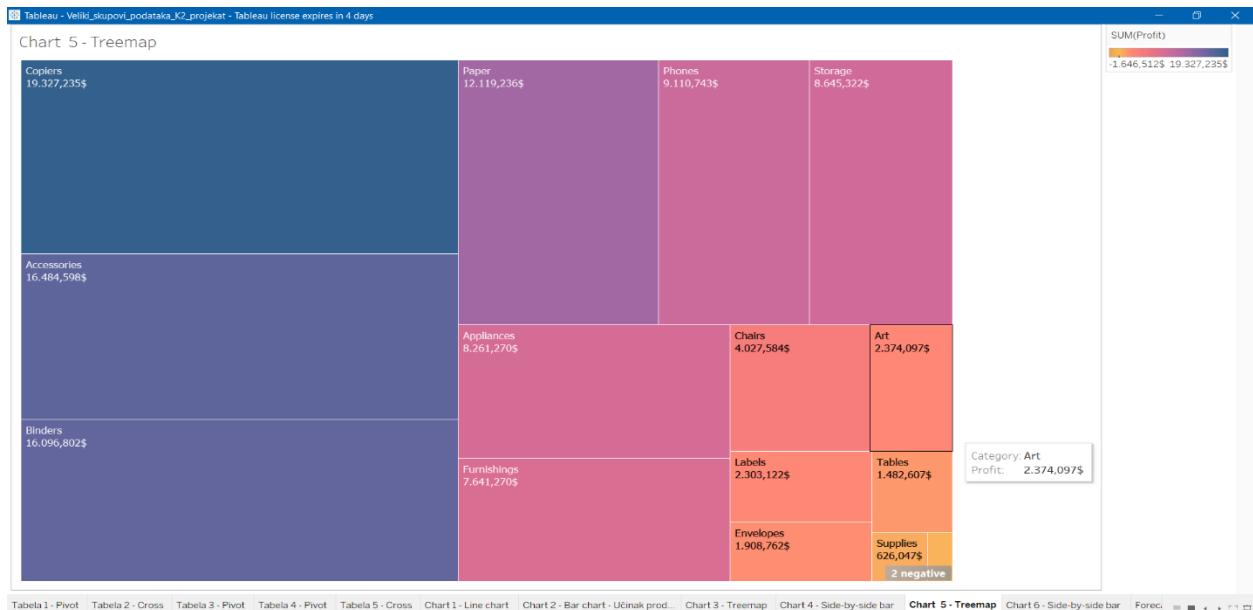
Četvrti grafik, nazvan Chart 4, je vrsta side-by-side grafika i predstavlja prosečan rok isporuke proizvoda za svaku državu isporuke, prikazan u satima. Najveće trajanje

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



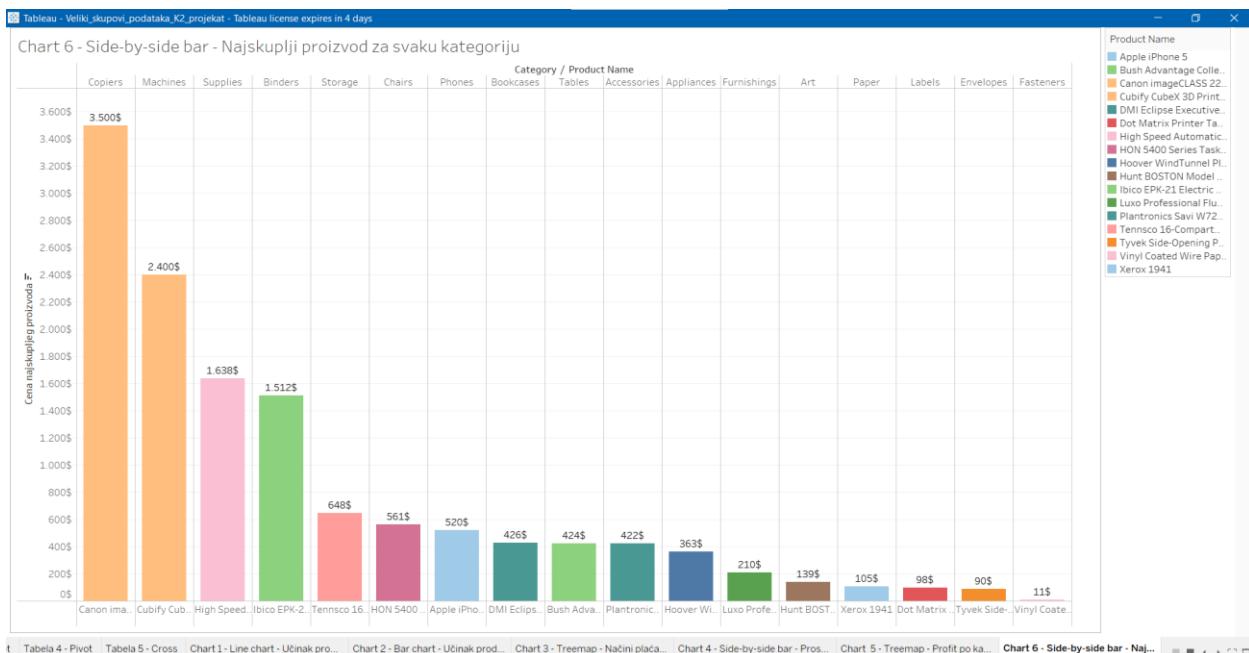
Slika 13 – Chart 4 prikazuje prosečno vreme trajanja isporuke za svaku državu

Peta 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\$.

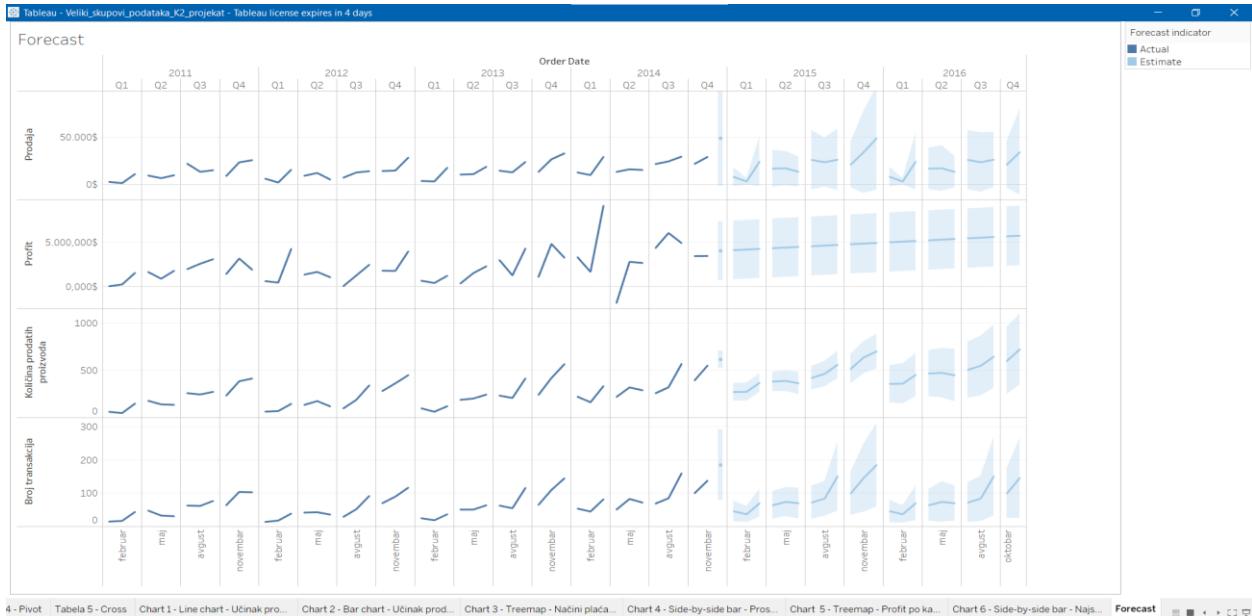


Slika 15 – Najskuplji proizvod iz svake kategorije i njegova cena

## 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

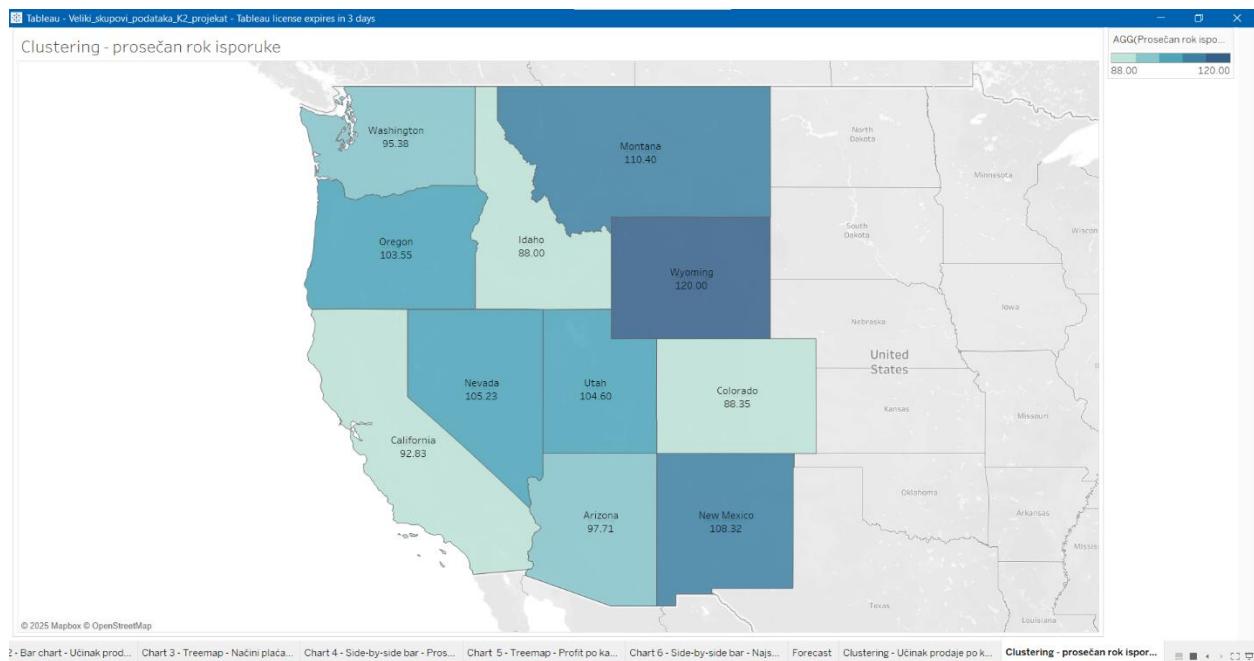
sekciju pozicioniramo dimenzije čiju vrednosti želimo da predviđamo, koje su u slučaju našeg projekta Profit, Prodaja, Količina prodatih proizvoda i Broj transakcija.U Analytics sekciji biramo opciju Forecast, tako što je drag-and-drop akcijom miša pozicioniramo na radnu površinu.Tableau automatski vrši 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

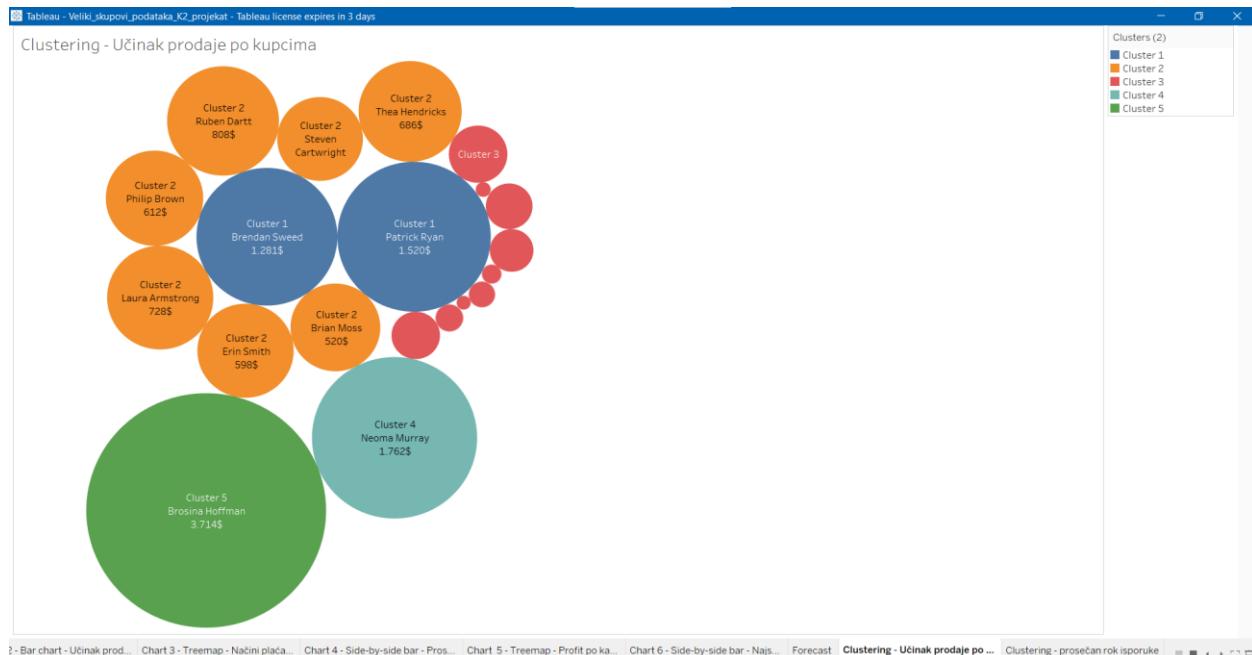
Kreiranje klastera je proces deljenja podataka na male grupe (klastere), gde su podaci unutar klastera grupisani po sličnim i bliskim vrednostima.Proces kreiranja klastera u Tableau softveru je sličan proceduri predviđanja(Forecast), gde se u sekcijama Columns i Rows definišu dimenzije podataka koje 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 klastere(grupe) sličnih podataka i obeleženi različitim bojama.

Prvo grupisanje podataka u projektu započeto je pozicioniranjem kolone *State* u Rows sekiju 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. Ikoniku 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).