

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

ПРЕДМЕТ: Складишта података

**Пројектовање складишта података за сајт који се
бави интернет трговином**

Ментори

проф. др Иван Луковић
др Сандро Радовановић

Студенти

Стеван Нешић 2022/3159
Анђела Вучковић 2022/3164

Београд, 2022. године

САДРЖАЈ

1. Увод.....	4
2. Циљеви, план и задаци развоја пројекта	5
2.1 Извор података	5
2.2 Истраживачка питања	6
2.3 ОЛТП шема за избрани пословни процес	7
2.4. Анализа и спецификација димензија и мера	8
3. Шема складишта података.....	12
3.1. Димензија Производ	12
3.2. Димензија Начин плаћања.....	13
3.3. Димензија Радња	13
3.4. Димензија Време	14
3.5. Табела чињеница (Поруџбина)	15
4. ЕТЛ процес.....	15
4.1. ЕТЛ пакет за попуњавање димензије Производ	17
4.2. ЕТЛ пакет за попуњавање димензије Начин плаћања.....	17
4.3. ЕТЛ пакет за попуњавање димензије Време	18
4.4. ЕТЛ пакет за попуњавање табеле чињеница (Поруџбина)	19
4.5. Пакет секвенцијалног извршавања (Execute)	20
5. Генерисање извештаја и одговори на постављена истраживачка питања	21
5.1. Прво истраживачко питање.....	21
5.2. Друго истраживачко питање	22
5.3. Треће истраживачко питање	23
5.4. Четврто истраживачко питање.....	24
6. Закључак.....	26
Литература.....	27

Попис слика

Слика 1 Шема.....	7
Слика 2 Графовски модел складишта података.....	8
Слика 3 Шема Димензије Производ	13
Слика 4 Шема Димензија Начин плаћања.....	13
Слика 5 Шема Димензија Радња	14
Слика 6 Шема Димензија Време	14
Слика 7 Шема Табеле чињенице (поруџбине).....	15
Слика 8 ЕТЛ	16
Слика 9 ЕТЛ Димензија Производ.....	17
Слика 10 ЕТЛ Димензија Начин плаћања	18
Слика 11 ЕТЛ Димензија Време.....	19
Слика 12 ЕТЛ Чињеничне табеле.....	20
Слика 13 ЕТЛ Секвенцијалног извршавања	21
Слика 14 Прво истраживачко питање.....	22
Слика 15 Друго истраживачко питање	23
Слика 16 Треће истраживачко питање	24
Слика 17 Четврто истраживачко питање.....	25

Попис табела

Табела 1 Истраживачка питања.....	6
Табела 2 Бус матрикс.....	7
Табела 3 Димензија Производ	9
Табела 4 Димензија Радња	10
Табела 5 Димензија Начин плаћања	10
Табела 6 Димензија Време	11
Табела 7 Мере.....	12

1. Увод

У 21. веку немогуће је замислити пословање и пружање услуга без савремених информационих система у чијим се позадинама налазе базе и складишта података, зависно од величине компаније.

Складишта података (енг. data warehouses) су специјализовани софтверски системи који се користе за складишење, управљање и анализу великих количина података. Они се често користе у пословним окружењима где се прикупљају подаци из различитих извора, као што су базе података, веб сајтови и сензори, а затим се складиште у једном месту за каснију анализу. Складишта података често се користе за бизнис интелигенцију и одлучивање на основу података, као и за друге аналитичке и истраживачке задатке. Они могу да помогну у откривању трендова, предвиђању будућих догађаја и оптимизацији пословних процеса. ("The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" by Ralph Kimball and Margy Ross)

Коришћењем различитих технологија у неколико етапа смо од класичних извора података преузетих као ексел табеле са интернета дошли до формирања складишта података кроз деномарнизацију табела и ЕТЛ процес.

ЕТЛ процес (енг. Extract, Transform, Load) је процес прикупљања података из различитих извора, их трансформације у жељени формат и учитавања у циљано складиште података. Овај процес је од кључног значаја за успех складишта података, јер омогућава ефикасну и тачну анализу података. Прва фаза процеса је екстракција (Extract) података из различитих извора, као што су базе података, текстуални фајлови и веб сајтови. Након екстракције, подаци се трансформишу (Transform) у облик који је прихватљив за складиште података. Овај процес укључује чишћење, уједностављивање и обогаћивање података. Коначно, учитавају се (Load) у циљано складиште података. ("Data Integration Blueprint and Modeling: Techniques for a Scalable and Sustainable Architecture" by Anthony David Giordano)

У другом делу рада главна тежња је била дати ваљане одговоре на постављена истраживачка питања као и поткрепити дате одговоре чињеницама, тј. визуелним ефектима. То је омогућено коришћењем Power BI платформе. Повер БИ (енг. Power BI) је бизнис интелигенција и аналитички сервис који је развио Мајкрософт. Он омогућава компанијама да прикупљају, анализирају и деле податке на ефикасан и интерактиван начин. Повер БИ укључује визуелизацију података, бизнис кључне показатеље (KPI-је), машинско учење и друге напредне алате за анализу података.

2. Циљеви, план и задаци развоја пројекта

2.1 Извор података

За потребе овог пројекта коришћена је база података интернет продаје бразилског сајта *Olist*, платформа која повезује мала и средња предузећа са купцима и која је специјализована у областима логистике и капитала. Трговина се променила, па платформе као што су *Shopify*, *Amazon*, *Alibaba* и остале су све релевантније за економије многих земаља. Овај покрет је подржао мала предузећа да се крећу у богатом и фрагментисаном систему.

Olist је број 1 за мала и средња предузећа који се сада шири на свет. Компанија је почела са једним језгром који повезује трговце са тржиштима и еволуирала је у комплетан систем интегрисаних производа у 3 димензије:

1. Трговина:
 - *Olist Store* је водеће решење за продају на тржиштима;
 - *Olist Shops* је наше решење за е-трговину, прво мобилно, изворно повезано са друштвеним медијима, присутно у 180 земаља.
2. Логистика: *Olist Pack* је водећи провајдер логистике и мреже за испуњење заснован на облаку који послује у Бразилу.
3. Капитал: *Olist Credit* и *Olist Pay*

У извору података се могу препознати пословни сценарији везани за различите услуге које компанија пружа, попут продаје, куповине, производа које фирме нуде, класификације и квалитета производа на основу искустава купаца. Овај рад се бави продајом, куповином и провером кључних фактора за успешно пословање компанија за посматрани временски период (конкретно у раду су коришћени подаци који покривају период од септембра 2016. године до октобра 2018. године).

Како је у оквиру рада препозната интернет продаја, односно препродаја, изабрани фокус је на сарадњи са фирмама-продавцима и фирмама или физичким лицима-купцима.

Подаци су прикупљени из следећег извора:

www.kaggle.com/datasets/olistbr/brazilian-ecommerce

2.2 Истраживачка питања

Фокус овог пројекта је на сагледавању ефикасности постојећих решења, формирање складишта податка зарад лакших приказа коришћења и очувања меморије као и визуелизација података ради помоћи доносиоцима одлука у унапређењу постојећих и развијању нових решења.

Осим дефинисања пословног процеса, потребно је имати и додатна објашњења, која су успостављена преко истраживачких питања. Питања су произвољно изабрана у складу са крајњим циљем, а то је како унапредити услуге куповине и продаје, представљено у табели 1.

Истраживачка питања	Опис
На који начин регија у коју се производ шаље утиче на број поруџбина и цену доставе?	На који начин насељеност и географски положај регије утичу на поручивање производа? Да ли је цена доставе зависна у односу на географски положај или у односу на број достава за одређену регију у неком временском периоду?
Која је зависност куповине у односу на време по кварталима и данима у недељи?	Да ли су људи склонији куповини у хладнијим или топлијим периодима године? Да ли је реална ситуација да више производа поручују онлајн током радне недеље него викендом због могућности одласка у радњу?
Какав је утицај начина плаћања у односу на број извршених плаћања и број рата?	У којој мери могућност плаћања на рате утиче на број извршених поруџбина? Какав је однос броја извршених поруџбина ако је плаћање до 10 рата у односу на маскималне 24 рате?
Како просечна цена производа по категорији утиче на поручивање ?	Која је то категорија која се више поручује у односу на остале и да ли њена просечна цена има утицаја?

Табела 1 Истраживачка питања

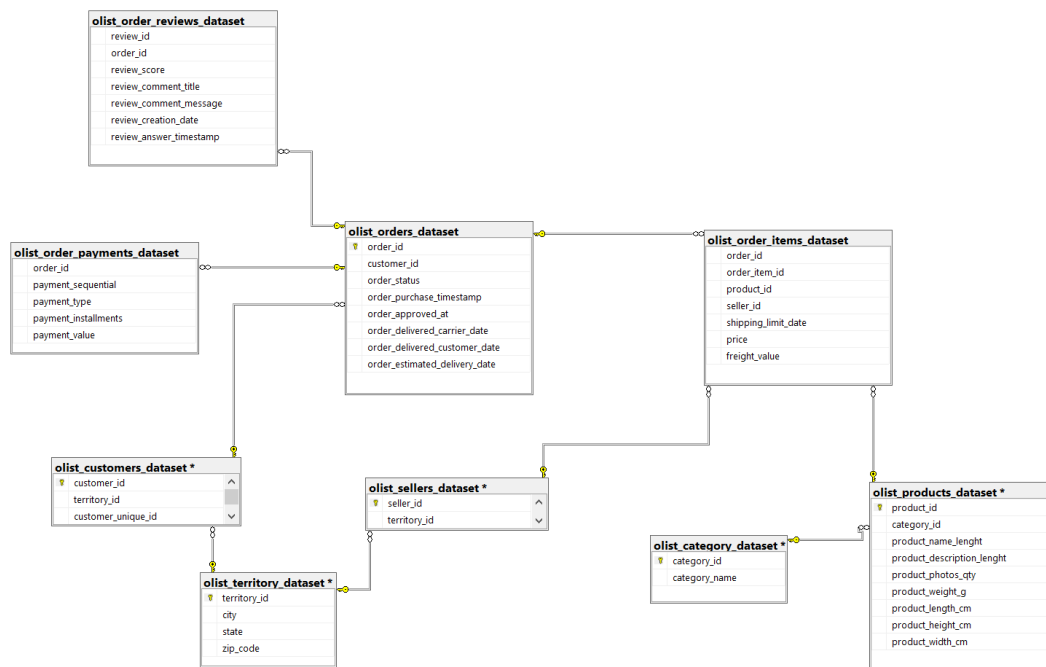
Пословна матрица (енг. *Bus matrix*) приказује односе између табела димензија и различитих процеса пословања, тако да се лако идентификује који пословни процеси имају исте димензије, а у којима се разликују. На основу добијених информација из матрице, могуће је проценити које табеле се могу комбиновати у један извештај. На Табели 2 је дат пример једне пословне матрице ове компаније.

Тема/Процес пословања		
Димензије	Утицај регије на доставу	Утицај плаћања на рате на продају
Производ	+	+
Радња	+	+
Време	+	+
Начин плаћања		+

Табела 2 Бус матрикс

2.3 ОЛТП шема за изабрани пословни процес

На основу OLTP извора података приказаног на слици 1, потребно је издвојити табеле које се тичу пословног процеса, а које су неопходне за настајање димензија и табела чињеница складишта података.



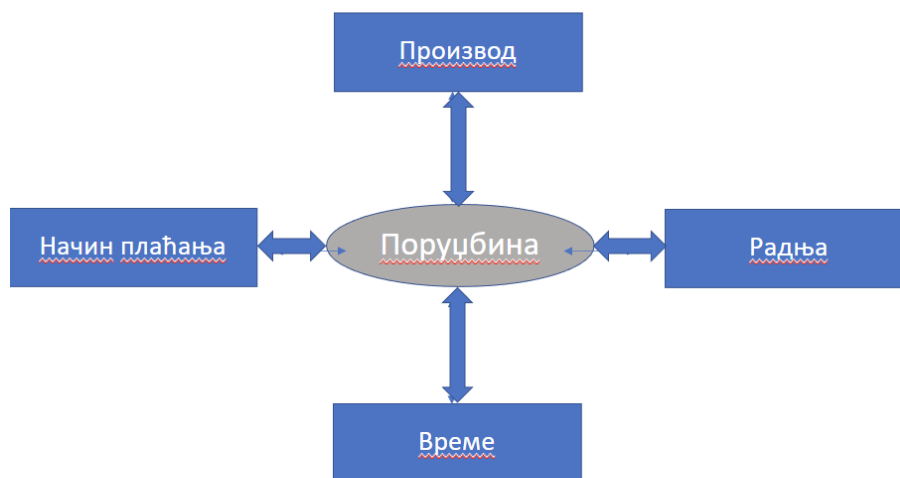
Слика 1 Шема

На основу претходно постављених истраживачких питања и шеме извора података, уочавамо димензије које ће бити формиране. У контексту продаје по кварталима и данима у недељи, где је јасно постављена реч о времену и димензији која ће представљати време. Димензија производ, као предмет продаје и предмет куповине заједно са својим категоријама, представља димензију од значаја за пословни процес. Обзиром да се ради о бразилској платформи за е-куповину, територија нам је ограничена на државу Бразил, која је подељена на федеративне републике, у тексту регије, а затим на радње, као једна од димензија. Услед великих разлика између богатих и сиромашних, јављају се различите потребе за начинима плаћања у односу на имовинско стање, као што је плаћање на рате или у кешу. Због тога начин плаћања представља четврту димензију.

2.4. Анализа и спецификација димензија и мера

Након избора табела на основу којих се формирају димензије и чињеничка табела, неопходно је дефинисати атрибуте за сваку димензију и чињеницу који ће бити потребни за одговарање на истраживачка питања и решавање пословног проблема. Већина атрибута се узима у оригиналном облику из извора података, али неки атрибути морају бити направљени самостално, на пример коришћењем скуповних функција или других начина израчунавања из постојећих атрибута. Зато је потребно извршити анализу и спецификацију корисничких захтева и извести атрибуте из сваке табеле који ће најбоље одговарати на постављена истраживачка питања, у оригиналном или изведеном облику.

На Слици 2 је приказан графовски модел будућег складишта података.



Слика 2 Графовски модел складишта података

Димензије су:

- Производ
- Радња
- Време
- Начин плаћања

Табела чињеница ће да указује на наруџбине препродаваца и мере које ће да садржи су:

- Цена производа
- Трошкови испоруке
- Оцена производа

Од којих табела из извора података настају нове табеле у складишту података је већ било речи, сада је потребно извршити спецификацију димензија, а потом и мера. У оквиру Табеле 3 је приказана спецификација димензије производа.

Грануларност	Назив производа	
Хијерархија	Назив производа, категорија производа	
Атрибути и домени	Атрибути	Домени
	ProductID	String
	CategoryName	{ Art, Music, Toys, Fashion shoes... }

Табела 3 Димензија Производ

Приликом изградње димензије производ уочљива је хијејархија, тако да категорија производа може имати један или више производа. У складишту података, два различита нивоа хијерархије ће бити приказана једном шемом релације, односно биће извршена денормализација, а жељени ниво грануларности је сам ИД производа. Наредна димензија је димензија радње и у оквиру Табеле 4 је приказана спецификација ове димензије.

Грануларност	Радња
---------------------	-------

Хијерархија	Радња, Град, Регија	
Атрибути и домени	Атрибути	Домени
	State	{SP, PR, BA, SC ...}
	City	{Sao Paolo, Rio de Janeiro, Porto Seguro ...}
	SellerID	String

Табела 4 Димензија Радња

И у овом случају постоји хијерархија, тако да важи да једна регија може имати један или више градова, а да један град обухвата једну или више радњи. Као и у претходном случају, тако и овде, биће извршена денормализација и сви нивои хијерархије ће бити приказани једном шемом релације.

Трећа димензија је начин плаћања и у оквиру Табеле 5 је приказана спецификација ове димензије.

Грануларност	Начин плаћања	
Хијерархија	Један ниво хијерархије	
Атрибути и домени	Атрибути	Домени
	Payment Type	{Boleto, Credit card, Debit card, Voucher}
	Payment Installments	{1,2,3,4,5,6,7,8...,23,24 }

Табела 5 Димензија Начин плаћања

За разлику од претходних димензија, у овом случају не постоји никаква хијерархија, односно реч је о једном нивоу хијерархије. Представљају се четири различита типа плаћања, што уједно представља и дефинисану грануларност, као и број рата за плаћање.

Нарочита важност димензије времена је у томе што се она разликује од осталих димензија по начину настанка. За разлику од осталих димензија које настају на основу шема релација ОЛТП извора података, димензију времена је потребно формирати самостално на основу неког или неких датумских атрибута. У конкретном случају, димензија времена је креирана на основу колоне OrderDate у табели наруџбеница, која представља датум поруџбине. Спецификација димензије времена може бити представљена у облику табеле, као што је приказано у Табели 6.

Грануларност	Дан	
Хијерархија	Дан, седмица, месец, квартал, година	
Атрибути и домени	Атрибути	Домени
	Day	{1,2,3,4,5,6,7}
	Month	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}
	Quarter	{1, 2, 3, 4}
	Year	{2016, 2017, 2018}

Табела 6 Димензија Време

Потребно је анализирати и спецификовати мере табеле чињеница, при чему ће неке мере бити узете у оригиналном облику из извора, док ће друге мере бити изведене на основу претходно узетих мера. Спецификација свих мера које остају у табели чињеница приказана у Табели 7.

	Назив мере		
	Цена производа	Трошкови испоруке	Оцена производа
Димензије	Производ, Време, Радња, Начин плаћања	Производ, Време, Радња,	Производ, Радња
Домен и јединица мере	Бразилски реал	Бразилски реал	[1, 2, 3, 4, 5]
Извор података	ОЛТП база података, табела са евиденцијом свих извршених трансакција препродаје за целокупни пњриод посматрања		

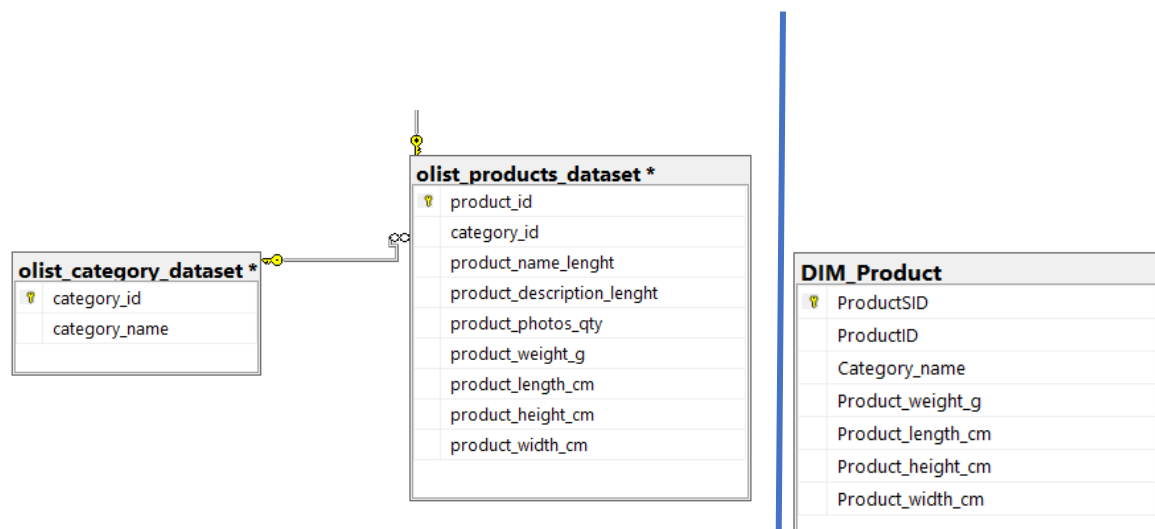
Табела 7 Мере

3. Шема складишта података

На основу ОЛТП шеме и пословних циљева направљене су табеле димензија и чињеница. Идентификоване су четири табеле димензија: DimDate, DimProduct, DimPayment и DimSeller. Како је у ОЛТП шеми података уочена хијерархија: ProductID -> CategoryName и SellerID -> City -> State, одлучено је да се изврши комплетна денормализација где се сви нивои хијерархије представљају једном шемом релација како би се добила жељена звездаста шема (енг. Star Schema), а о чему је у претходном поглављу било речи. На овај начин се обезбеђују боље перформансе приликом извештавања.

3.1. Димензија Производ

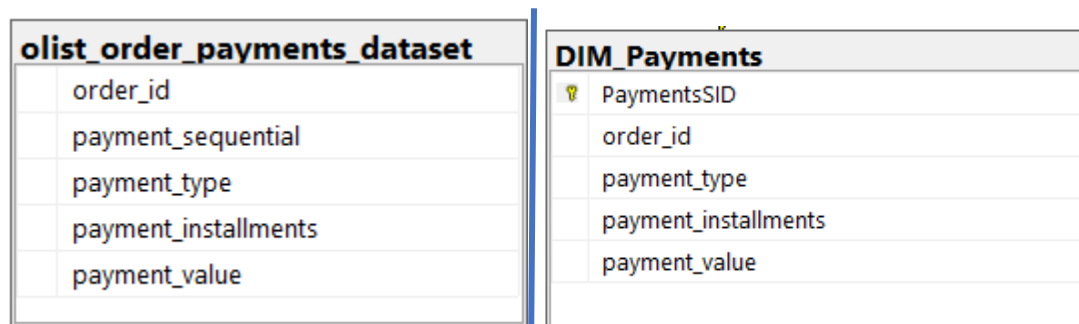
Димензија Производ (енг. DimProduct) је добијена спајањем табела *olist_category_dataset* и *olist_products_dataset* преко атрибута *categoryID*. Табела DimProduct садржи податке о самим производима: идентификациони број производа (енг. ProductID), назив категорије (енг. CategoryName), тежину, ширину, дужину и висину, као значајну информацију за доставу производа. На Слици 3 се може видети начин на који је формирана табела DimProduct.



Слика 3 Шема Димензије Производ

3.2. Димензија Начин плаћања

Димензија Начин плаћања (енг. *DimPayment*) има један ниво хијерархије и садржи податке који су везани за начин плаћања (енг. *Payment_type*), број рата(енг. *Payment_installments*) и износ плаћања(енг. *Payment_value*) (Слика 4).



Слика 4 Шема Димензија Начин плаћања

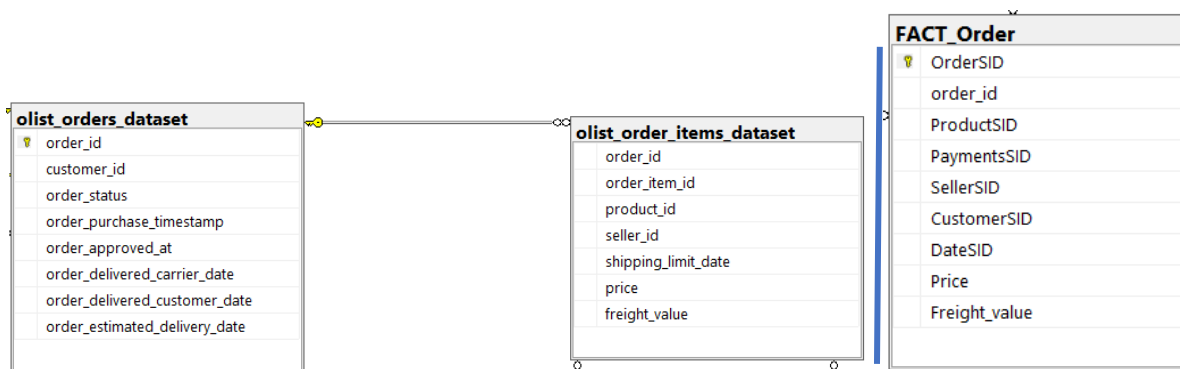
3.3. Димензија Радња

Димензија Радња (енг. *DimSeller*) чува податке који су везани за географски простор, и то: идентификациони број продавца, односно радње (енг. *SellerID*), град (енг. *City*) и регион (енг. *State*). Ова табела димензија се добија спајањем табела *olist_territory_dataset* и *olist_seller_dataset* на основу атрибута TerritoryID (Слика 5).

3.5. Табела чињеница (Поруџбина)

Поред табела димензија уочена је једна табела чињеница: *Fact_Order*, где се налазе атрибути везани поруџбину: број поруџбине (енг. *OrderID*), цена производа (енг. *Price*) и цена доставе (енг. *Freight_value*)

Како табела чињеница представља централну табелу у звездастој шеми, која референцира табеле димензија, постоје и четири додатна атрибута који представљају спољне кључеве ка овим табелама (*DateSID*, *PaymentSID*, *ProductSID*, *SellerSID*). На Слици 7 се може видети начин креирања табеле чињеница. Као и код табела димензија, и у табели чињеница је креиран вештачки примарни кључ *OrderSID*.



Слика 7 Шема Табеле чињенице (поруџбине)

4. ЕТЛ процес

Подаци у систему складишта података се учитавају након што се изврше све операције ЕТЛ процеса. ЕТЛ процес представља процес складиштења података који се користи за издвајање и извлачење података из различитих извора података, потом за трансформацију података и на крају, пуњење самог складишта подацима. Дакле, ЕТЛ процес се састоји од три функције:

- Екстракција,
- Трансформација и
- Учитавање.

Екстракција података представља процес читања података из различитих извора података. Сви подаци добијени из оваквих извора претварају се у исти формат складиштења података и прослеђују се у наредни корак, спровођење трансформације.

Трансформација података је поступак који захтева највише времена у ЕТЛ процесу. Подаци преузети из различитих извора се не могу користити у изворном облику. Потребно је извршити валидацију, прочишћавање, интеграцију и временско означавање података.

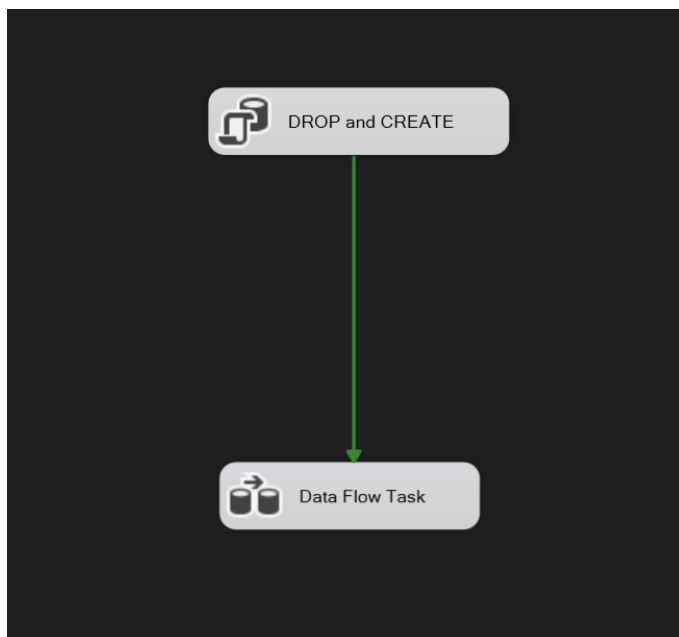
Када су подаци трансформисани и када су досегли одређени ниво квалитета, прелази се на поступак иницијалног пуњења складишта података, што представља последњи корак у ЕТЛ процесу.

За реализацију ЕТЛ процеса коришћени су SQL сервер интеграциони сервис (енг. SQL Server Integration Services - SSIS) у Visual Studio окружењу. SSIS је платформа за изградњу решења за интеграцију података високих перформанси, укључујући пакете за екстракцију, трансформацију и учитавање (ЕТЛ) у складиште података. Представља основни ЕТЛ алат који је део Microsoft - овог пакета Business Intelligence Suite (Knight et al., 2012).

У сваком од засебних пакета је дефинисан процес формирања табела димензија и табеле чињеница. Затим је у новом пакету креиран контејнер који омогућава секвенцијално извршење претходно дефинисаних пакета.

За учитавање података из ОЛТП базе података коришћен је ODBC Source алат. Приликом пуњења података коришћен је ODBC Destination. За потребе спајања табела употребљен је Lookup алат који врши спајање на основу примарног/спољног кључа. Замена недостајућих вредности, преименовање 20 вредности, креирање изведених атрибута ради се преко алата Derived Column. Елиминисање дуплих вредности је извршено уз помоћ Sort алата.

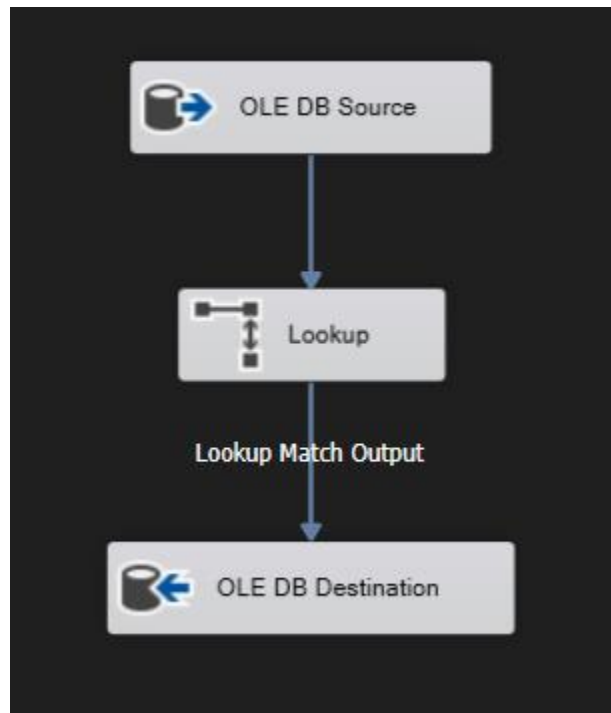
Извршавањем сваког ЕТЛ процеса, на почетку се прво дешава брисање претходно унетих података, како би се избегла редувантност, као и саме табеле, а затим се креира иста табела и поново пуни подацима (Слика 8).



Слика 8 ЕТЛ

4.1. ЕТЛ пакет за попуњавање димензије Производ

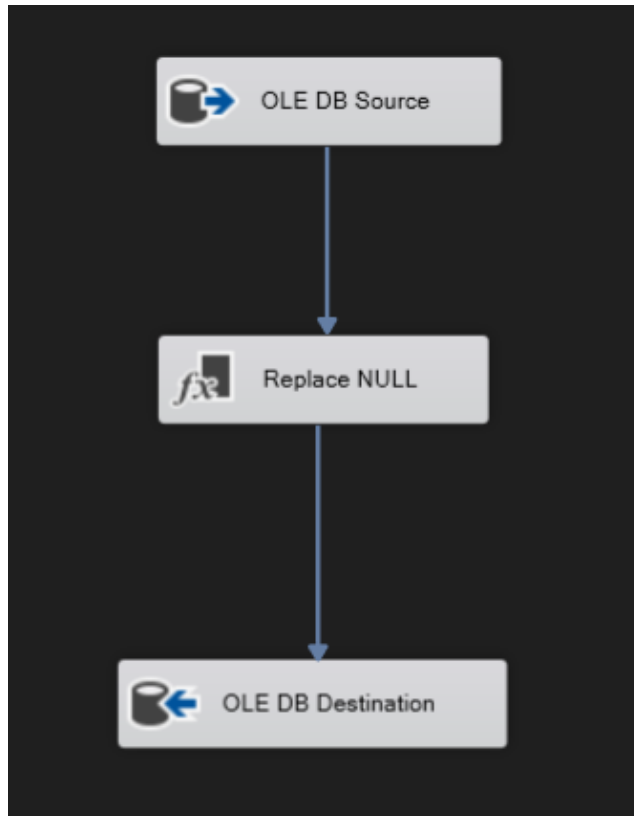
Прво су уčitане потребне колоне из табеле Производ. Затим је помоћу Lookup-а извршено спајање табеле Производ са табелом категорија производа. Такође, у оквиру Lookup-а су уčitане потребне колоне из табеле категорија производа. Производи који имају дефинисану поткатегорију су спојени са табелом Категорија производа и урађена је селекција колоне из ове табеле. Затим је извршена унија претхоног излаза и производа који немају дефинисану поткатегорију (Слика 9).



Слика 9 ЕТЛ Димензија Производ

4.2. ЕТЛ пакет за попуњавање димензије Начин плаћања

Обзиром да је у самом извору података дозвољена опција *Allow Nulls*, односно да су дозвољене нула вредности у атрибутима *Payment_installments* и *Payment_type*, одабрана је опција замене нултих вредности стринговима(*Replace NULL*) : “Не знамо како је плаћено” и “Не знамо број рата”.(Слика 10)

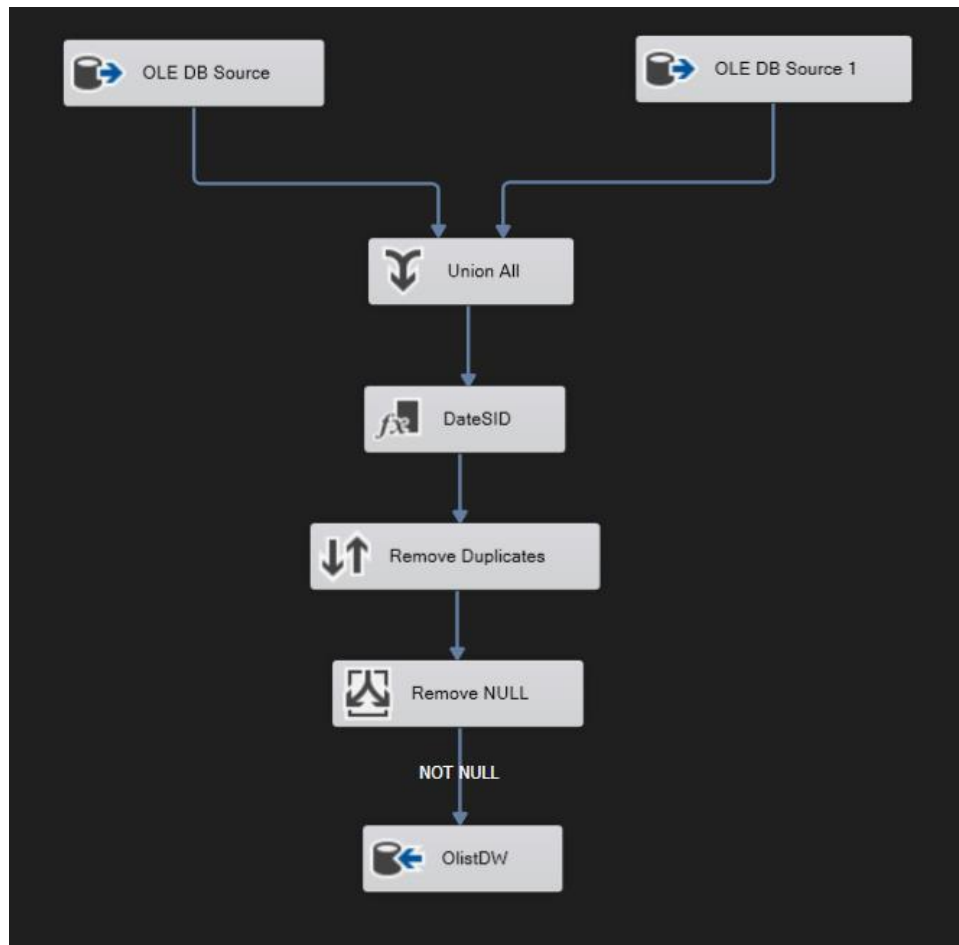


Слика 10 ЕТЛ Димензија Начин плаћања

4.3. ЕТЛ пакет за попуњавање димензије Време

Поступак креирања Execute SQL Task-а је исти као и у претходим пакетима. Data Flow Task почиње спајањем колона *Order_approved_at* и *Order_delivery_date* из табеле *olist_order_dataset*, зарад добијања потребних података, као и рачунања броја дана за доставу.

Пошто се у једном дану изврши више наруџбина, овако креирана табела ће имати више истих вредности. Садржаће 99,941 датума, управо онолико колико има поруџбина. Како бисмо овај проблем превазишли потребно је елиминисати поновљене вредности за атрибут *DateSID* како би он постао јединствени идентификатор. Након овог поступка добијамо табелу која има 678 редова (датума). Као и у претходним случајевима овај ЕТЛ процес се завршава пуњењем складишта података у димензију Време (Слика 11).



Слика 11 ЕТЛ Димензија Време

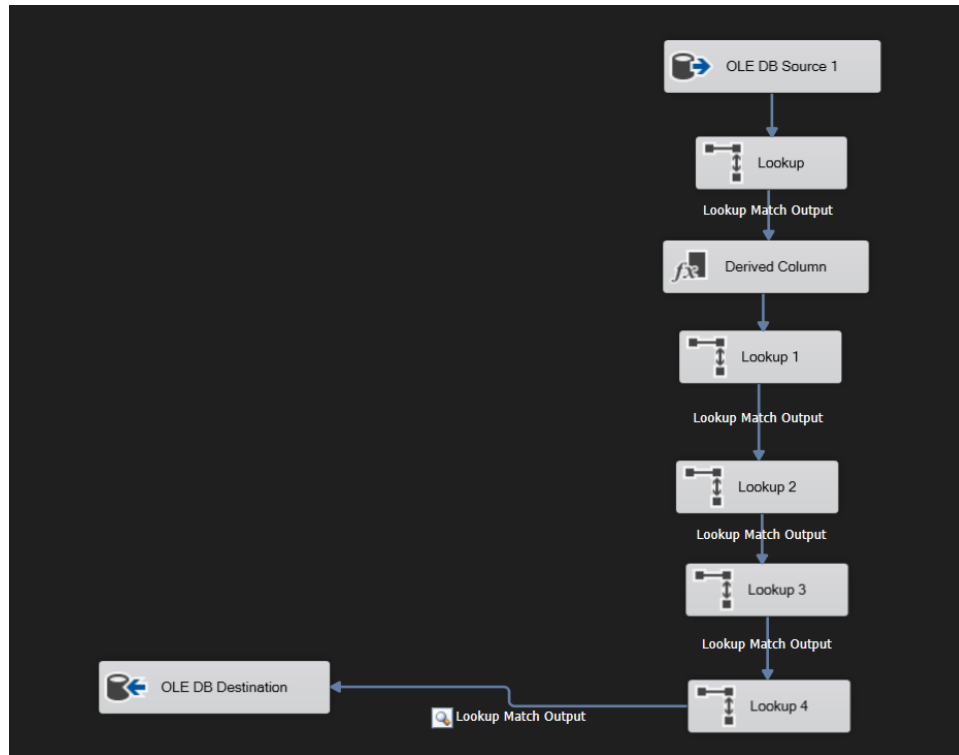
4.4. ЕТЛ пакет за попуњавање табеле чињеница (Поруџбина)

Execute SQL Task за табелу *FactOrder* се мало разликује од претходних пакета. На почетку се проверава да ли постоји табела *FactOrder*. У случају да табела постоји врши се брисање исте. Након тога се табела креира. Након креирања табеле потребно је поставити сва ограничења за спољне кључеве према свим табелама димензија.

Flow Task почиње читавањем потребних колона из табеле *Olist_orders_items_dataset*. Уочено је да ставке производа нису нумерисане према наруџбини, односно не може се јасно утврдити редни број ставке за конкретну наруџбину. Приликом читавања података, уз помоћ SQL упита, извршено је нумерисање ставки на основу наруџбина.

Након тога, врши се повезивање са табелом *Olist_orders_dataset*. Када су табеле повезане, потребно је извршити повезивање са табелама димензија. Битно је напоменути да се повезивање табела врши на основу примарних кључева из ОЛТП базе података, а не на основу новодефинисаних вештачких кључева, с тим што ће се вредност тих кључева чувати

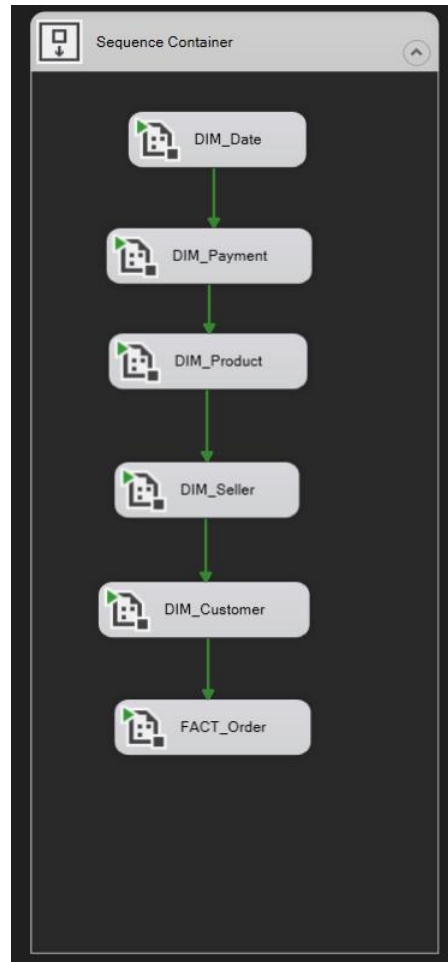
у табели чињеница. До смањења броја торки долази након спајања са осталим табелама. Као што је претходно напоменуто, у овој табели се чувају само купци који имају продавницу, односно који су препродавци. (Слика 12).



Слика 12 ЕТЛ Чињеничне табеле

4.5. Пакет секвенцијалног извршавања (Execute)

Да би било сигурно да ће се складиште података правилно попунити без грешака, потребно је утврдити редослед извршавања пакета који су претходно креирани. У пакету *FactOrder*, који користи Execute SQL Task, потребно је прво повезати табелу *FactOrder*, са табелама димензија, што значи да табеле димензија морају бити креиране пре извршавања пакета *FactOrder*. Потребно је извршити независне пакете, као што су пакети који се односе на креирање табела димензија, пре него што се изврши пакет који је зависан од свих претходно креираних пакета, тј. пакет који представља табелу чињеница. (Слика 13).



Слика 13 ЕТЛ Секвенцијалног извршавања

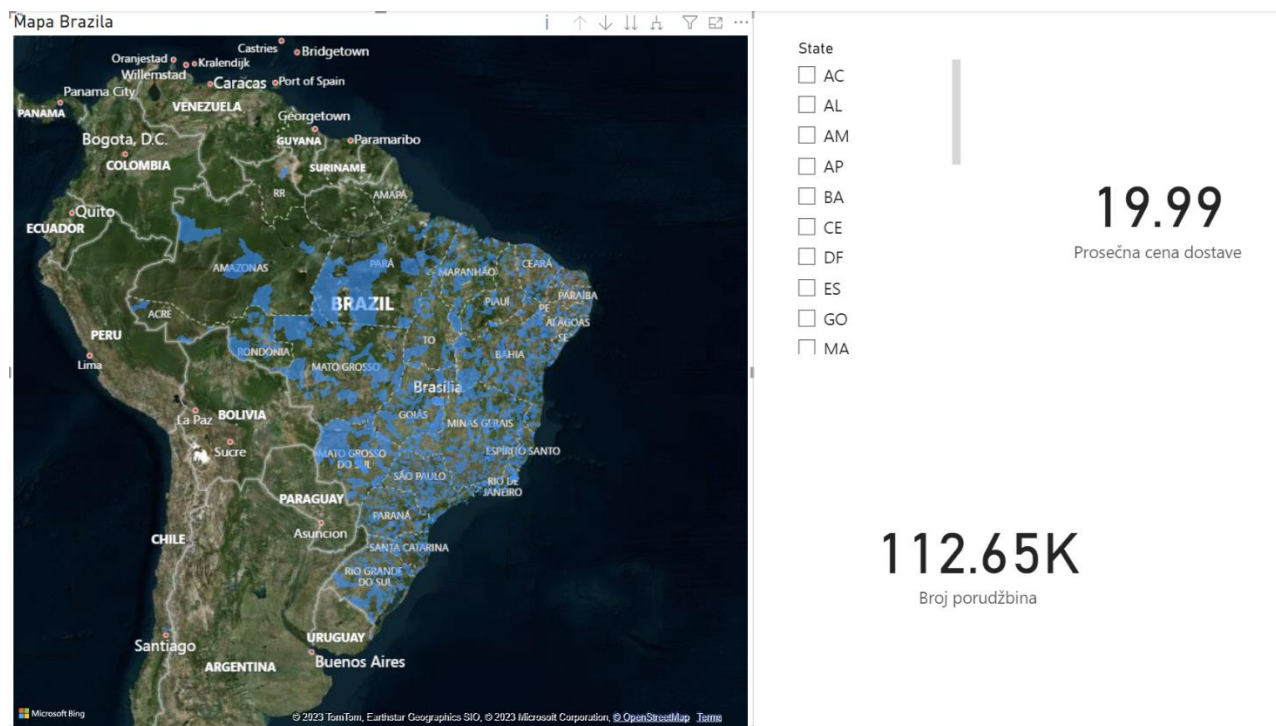
5. Генерисање извештаја и одговори на постављена истраживачка питања

Након успешног проласка кроз ЕТЛ процес и формирања складишта података уско повезаног са процесом пословања који се прати, створена је основа за генерисање извештаја који ће пружити подршку при одлучивању и доношењу одлука, а који ће, пре свега, визуелно и аналитички омогућити доношење одговорљ на жељена истраживачка питања.

5.1. Прво истраживачко питање

На који начин регија у коју се производ шаље утиче на број поруџбина и цену доставе?

На левој половини Сlike 14 приказана је заступљеност онлајн поручивања у регијама Бразила. Број поруџбина, у периоду за који имамо податке, је 112650, а просечна цена доставе је 19,99 BRL. Одабиром тачно одређене регије, понуђене у централном делу слике мапа се издефинише на тачно одређену регију и градове у њој, док се истовремено врши израчунавање броја поруџбина и просечна цена доставе за дату регију.



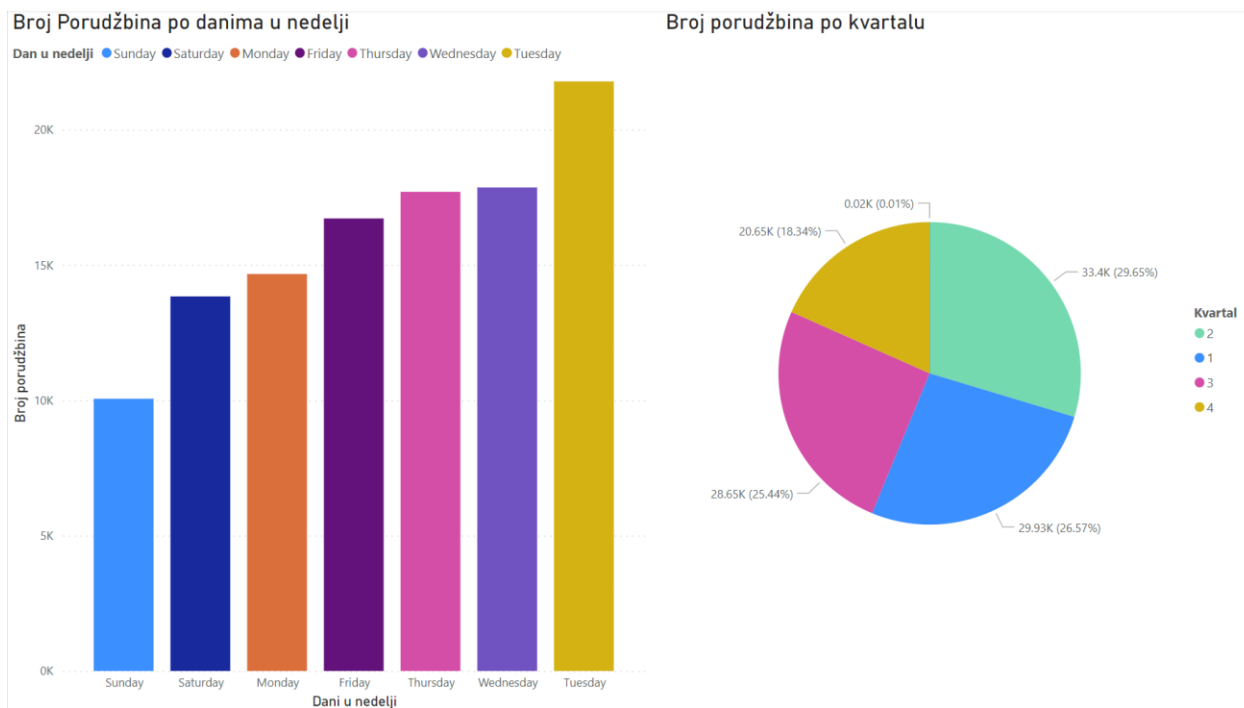
Слика 14 Прво истраживачко питање

Северни део Бразила познат по реци Амазон и највећој прашуми на свету, Амазонији, ретко је насељена област, самим тим је у датом периоду било најмање извршених поруџбина, као и највећа цена доставе. Лако је закључити да су број поруџбина и цена доставе обрнуто пропорционалне. Ако је број поруџбина за одређену регију мали, а трошак достављача је исти, купац на тај начин мора да надомести трошкове доставе. Закључили смо да је и број дана до достављања пакета видно већи због покушаја достављача да прикупи већи број поруџбина за ту област. Иста је и логика за гушће настањена подручја и већи број поруџбина.

5.2. Друго истраживачко питање

Која је зависност куповине у односу на време по кварталима и данима у недељи?

Зависност куповине у односу на дане у недељи огледа се у могућности или немогућности купца да, услед обавеза и свакодневних активности, физички оде у радњу и купи оно што му је потребно. Стога можемо приметити да је број онлајн поруџбина најмањи викендом, а највећи средином радне недеље, тј. уторком, средом и четвртком. (Слика 15)



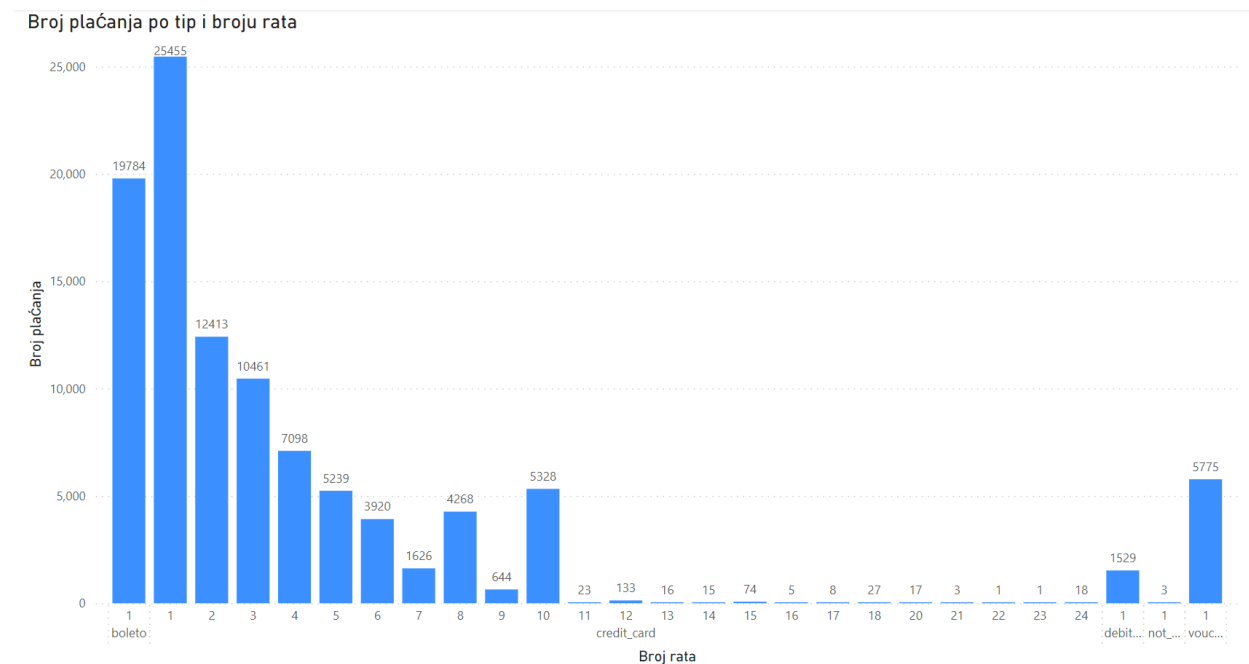
Слика 15 Друго истраживачко питање

Што се тиче прегледа по кварталима видимо да је највећи број онлајн поруџбина у пролеће, а најмањи у јесен, док је зими и лети приближно исти број поруџбина.

5.3. Треће истраживачко питање

Какав је утицај начина плаћања у односу на број извршених плаћања и број рата?

Могућност различитих начина плаћања, као и плаћања на рате је један од разлога зашто је онлајн куповина све популарнија. Као најзаступљенији вид плаћања видимо да је плаћање кредитном картицом и то са могућношћу плаћања до 10 рата, затим плаћање кредитном картицом одједном, док је плаћање кешом тек на трећем месту. Најмање заступљени вид плаћања представља кредитном картицом од 10 до 24 рате. Претпоставимо да разлог томе плаћање камате зависно од политике банке. (Слика 16)

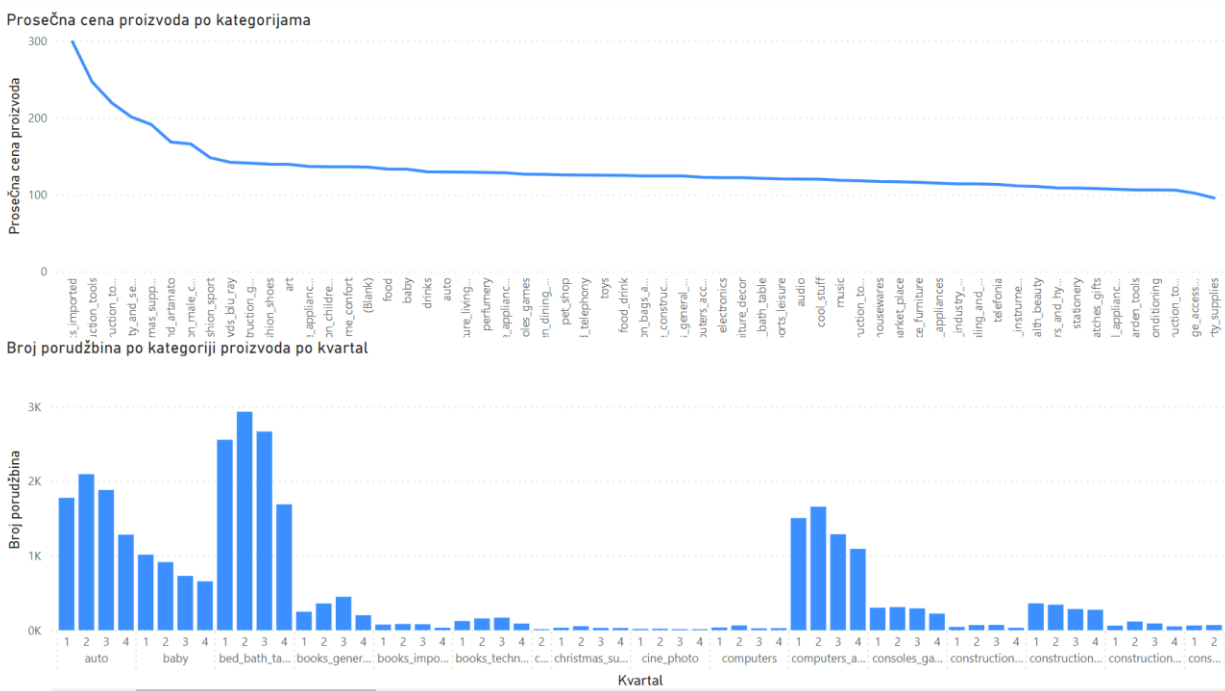


Слика 16 Треће истраживачко питање

5.4. Четврто истраживачко питање

Како просечна цена производа по категорији утиче на поручивање?

Обзиром на то да су у већини случајева мале разлике у ценама по категорија, што је изненађујућа чињеница, можемо закључити да просечна цена производа по категорији не утиче на поручивање, већ се оно обавља без икаквих правила сходно потребама купаца, што се може видети на слици 17.



Слика 17 Четврто истраживачко питање

Оно што можемо закључити је да су одређене категорије производа јако примамљиве за онлајн поручивање (*auto*, *baby*, *computer accessories...*), док су неке друге категорије веома непогодне за онлајн поручивање (*cine_photo*, *food*, *art...*).

6. Закључак

Анализом и одвајањем битних информација од оних мање битних, као и праћењем пословних процеса и циљева пословања, може се закључити много о самом предузећу, али и о могућностима његових унапређења. Како данас нагло расте број онлајн продавница, једини водич ка успеху и остварењу конкурентске предности јесте примена моћних концепата пословне интелигенције.

Проласком кроз претходна поглавља и анализом скупа података, који је коришћен за потребе овог рада, изведени су одређени закључци битни за пословање предузећа и валидно доношење одлука. На основу елабората циљева, планова и задатака пројекта, развијене су све неопходне етапе за развој пројекта. Из извора података издвојене су потребне шеме релације и атрибути за креирање складишта података. Детаљним проласком кроз ЕТЛ процес, креирана је свака од замишљених димензија података, а потом и табела чињеница са свим неопходним мерама за утврђивање успешности пословања предузећа, са акцентом на прибављање одговора на постављена истраживачка питања. Оформљена звездаста шема садржи четири димензије: *DimProduct*, *DimSeller*, *DimPayment* и *DimDate*, као и табелу чињеница *FactOrder*.

Циљ ове анализе је да на основу добијених и изведених резултата помогне доносиоцима одлука у одлучивању, као и да побољша однос продаваца према купцима и обрнуто, тако да обе стране буду задовољније.

Литература

- [1] "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" by Ralph Kimball and Margy Ross
- [2] "Building the Data Warehouse" by W.H. Inmon
- [3] "Data Integration Blueprint and Modeling: Techniques for a Scalable and Sustainable Architecture" by Anthony David Giordano
- [4] "ETL with SQL Server Integration Services" by Siva Harinath and Dandy Weyn
- [5] "Practical Power BI: Transform, Model, and Visualize Data" by Brett Powell