

Univerzita Pardubice
Fakulta ekonomicko-správní

RFM analýza
Bc. Veronika Nováková

Diplomová práce
2021

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2020/2021

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Veronika Nováková**
Osobní číslo: **E19628**
Studijní program: **N0413A050009 Ekonomika a management**
Studijní obor: **Ekonomika a management podniku**
Téma práce: **RFM analýza**
Zadávací katedra: **Ústav podnikové ekonomiky a managementu**

Zásady pro vypracování

Cíl: Návrh příkladu využití RFM analýzy na dostupných datech.

Osnova:

- definice základních pojmů ve zvolené oblasti,
- formulace problému,
- zpracování dat pomocí RFM analýzy,
- vyhodnocení výsledků a formulace závěrů.

Rozsah pracovní zprávy: **cca 50 stran**
Rozsah grafických prací:
Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam doporučené literatury:

- BERKA, Petr. *Dobývání znalostí z databází*. Praha: Academia, 2003. ISBN 80-200-1062-9.
BERRY, Michael J. A. *Data mining techniques: for marketing, sales, and customer relationship management*. Indianapolis: Wiley, 2004. ISBN 0-471-47064-3.
HAN, Jiawei a Micheline KAMBER. *Data mining: concepts and techniques*. 2nd ed. San Francisco: Morgan Kaufmann, c2006. Morgan Kaufmann series in data management systems. ISBN 1-55860-901-6.
HENDL, Jan. *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. Praha: Portál, 2004. ISBN 80-7178-820-1.
RUD, Olivia Parr, Ivo MAGERA a Milan DANĚK. *Data mining: praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníků (CRM)*. Praha: Computer Press, 2001. Databáze. Rychle a jistě. ISBN 80-7226-577-6.
WENDLER, Tilo a Sören GRÖTTRUP. *Data mining with SPSS modeler: theory, exercises and solutions*. Cham: Springer International Publishing, 2016. ISBN 978-3-319-28707-2.
Zdroje Internetu.

Vedoucí diplomové práce: **Ing. Miloslava Kašparová, Ph.D.**
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: **1. září 2020**
Termín odevzdání diplomové práce: **30. dubna 2021**

prof. Ing. Jan Stejskal, Ph.D. v.r.
děkan

L.S.

doc. Ing. Marcela Kožená, Ph.D. v.r.
vedoucí ústavu

V Pardubicích dne 1. září 2020

PROHLÁŠENÍ

Prohlašuji:

Práci s názvem RFM analýza jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnicí Univerzity Pardubice č. 7/2019 Pravidla pro odevzdávání, zveřejňování a formální úpravu závěrečných prací, ve znění pozdějších dodatků, bude práce zveřejněna prostřednictvím Digitální knihovny Univerzity Pardubice.

V Pardubicích dne 30.6.2021

Bc. Veronika Nováková

PODĚKOVÁNÍ:

Tímto bych ráda poděkovala vedoucí diplomové práce Ing. Miloslavě Kašparové, Ph. D. za její trpělivost, odbornou pomoc a cenné rady, které mi pomohly při zpracování této diplomové práce. Dále bych chtěla poděkovat mé rodině za trpělivost a podporu během celých studií.

ANOTACE

Cílem této diplomové práce je navrhnout příklad využití RFM analýzy na dostupných datech. Tato práce se zabývá definicí základních pojmů v této oblasti, formulací problému, zpracováním dat pomocí RFM analýzy a následné vyhodnocení výsledků a formulace závěrů.

KLÍČOVÁ SLOVA

RFM analýza, recency, frequency, monetary

TITLE

RFM Analysis

ANNOTATION

The goal of this diploma thesis is to propose an example of using RFM analysis on available data. This thesis is focused on the definition of basic concepts in this area, problem formulation, data processing using RFM analysis and subsequent evaluation of results and formulation of conclusions.

KEYWORDS

RFM analysis, recency, frequency, monetary

OBSAH

ÚVOD.....	10
1 DEFINICE ZÁKLADNÍCH POJMŮ VE ZVOLENÉ OBLASTI	11
1.1 RFM ANALÝZA.....	11
1.1.1 Bodování RFM	12
1.1.2 Kvantily zákazníků	12
1.1.3 Hodnocení kvantilů chování.....	13
1.1.4 Výhody RFM modelu	14
1.1.5 Nevýhody RFM modelu.....	15
1.1.6 Fáze zákazníků.....	15
1.2 URČENÍ RFM	16
1.2.1 Metoda vážení.....	16
1.2.2 RFM a shlukování.....	17
1.3 HODNOTA PRO ZÁKAZNÍKA	19
1.4 SEGMENTY ZÁKAZNÍKŮ	21
2 FORMULACE PROBLÉMU.....	24
3 ZPRACOVÁNÍ DAT POMOCÍ RFM ANALÝZY	29
3.1 IBM SPSS MODELER	29
3.2 PŘÍPRAVA DAT	30
3.3 RFM V PROSTŘEDÍ IBM SPSS MODELER.....	32
3.4 ZPRACOVÁNÍ DAT	33
3.4.1 RFM analýza – přístup 1	40
3.4.2 RFM analýza – přístup 2	43
3.4.3 RFM analýza – přístup 3	46
4 VYHODNOCENÍ VÝSLEDKŮ A FORMULACE ZÁVĚRŮ	51
4.1 VYHODNOCENÍ PŘÍSTUPU 2.....	51
4.2 VYHODNOCENÍ PŘÍSTUPU 3.....	54
ZÁVĚR.....	58
POUŽITÁ LITERATURA	60

SEZNAM ILUSTRACÍ A TABULEK

Obrázek 1 Tradiční a moderní zákaznický orientovaná struktura	21
Obrázek 2 Ukázka prvotních poskytnutých dat	25
Obrázek 3 Podíl měsíců na celkové hodnotě za dané období.....	26
Obrázek 4 Celková výše útrat v letech 2018-2020	27
Obrázek 5 Celková výše útrat v jednotlivých měsících	28
Obrázek 6 Stream pro přiřazení ID zákazníka.....	30
Obrázek 7 Ukázka dat s přidělením ID zákazníka	31
Obrázek 8 Prvotní datový audit	32
Obrázek 9 Stream RFM analýzy	33
Obrázek 10 Hodnoty R, F a M.....	34
Obrázek 11 Histogram hodnoty R.....	35
Obrázek 12 Histogram hodnoty F	36
Obrázek 13 Histogram hodnoty M.....	37
Obrázek 14 Krabicový graf hodnoty R	38
Obrázek 15 Krabicový graf hodnoty F	39
Obrázek 16 Krabicový graf hodnoty M.....	40
Obrázek 17 RFM skóre – přístup 1	42
Obrázek 18 3D vizualizace – přístup 1.....	43
Obrázek 19 RFM skóre – přístup 2	45
Obrázek 20 3D vizualizace – přístup 2.....	45
Obrázek 21 Datový audit.....	46
Obrázek 22 K-Means stream	47
Obrázek 23 Procentní znázornění klastrů	47
Obrázek 24 Průměrné hodnoty klastrů	48
Obrázek 25 3D vizualizace – přístup 3.....	49
Obrázek 26 Koeficient siluety	50
Obrázek 27 Zařazení jednotlivých ID do klastrů	50
Obrázek 28 Doplněný stream o distribuční graf – přístup 2.....	51
Obrázek 29 Ukázka části distribučního grafu s RFM skóre	52
Obrázek 30 Podíl jednotlivých segmentů	54
Obrázek 31 Doplněný stream o distribuční graf – přístup 3.....	55
Obrázek 32 Distribuční graf klastrů.....	55
Obrázek 33 Podíl jednotlivých fází zákazníků	57
Tabulka 1 Datový slovník prvotních dat	25

SEZNAM ZKRATEK A ZNAČEK

3D – trojrozměrný

CPV – hodnota vnímaná zákazníkem

F – četnost

ID – identifikace

M – peněžní hodnota

R – aktuálnost

RFM – RFM analýza

ÚVOD

Jako téma své závěrečné diplomové práce jsem si zvolila RFM analýzu. Důvod výběru právě tohoto tématu byla má neznalost této problematiky. Po krátkém nastudování toho, o čem samotná analýza je, jsem usoudila, že toto téma bude pro mě velmi zajímavé a ráda se o něm dozvím více.

V dnešní době existuje spousta marketingových analýz, které se zabývají segmentací zákazníků, některé jsou složitější, jiné naopak lehčí. V případě této analýzy se vychází ze samotného nákupního chování každého zákazníka a je používána v přímé marketingové segmentaci, zaobírajíc se marketingovým kontaktováním zákazníků. Podstata této analýzy je v celku jednoduchá, posuzují se zde tři nejzákladnější údaje každého nákupu, a to základní údaj o zákazníkovi, ve většině případů se jedná o jedinečné ID zákazníka, datum nákupu a hodnota nákupu. Výhodou RFM analýzy je pak stanovení, zda si daný zákazník zaslouží být znovu kontaktován a vynakládat na něj tak další náklady či nikoliv, jelikož už od něj třeba nemá společnost po dlouhou dobu žádnou odezvu formou uskutečnění nového nákupu.

Hlavním cílem mé práce je tedy vysvětlení základní problematiky RFM analýzy a návrh příkladu využití této analýzy na dostupných datech.

Diplomová práce je rozdělena do čtyř hlavních kapitol. První kapitola je věnována definici základních pojmů v oblasti RFM analýzy. Jedná se o samotné vysvětlení podstaty, výhod a nevýhod použití, výpočtu či určování RFM analýzy. Dále se tato kapitola zabývá hodnotou pro zákazníka, tj. co je vlastně pro zákazníka to hlavní, proč u společnosti nadále nakupuje. Poslední část této kapitoly je zaměřena na různé přístupy segmentace zákazníků v rámci této problematiky.

Ve druhé kapitole již dochází k formulaci problému neboli co bude touto analýzou řešeno. Tato kapitola pojednává o základním představení společnosti, jejich současné segmentaci zákazníků a dále se zabývá následnou vizualizací poskytnutých dat.

Třetí kapitola je již věnována představení softwaru IBM SPSS Modeler, v jehož prostředí dojde k provedení analýzy. Dále je v této kapitole zahrnuta potřebná příprava dat, zpracování dat a vytvoření tří různých přístupů uskutečnitelných právě v tomto softwaru v rámci RFM analýzy.

Čtvrtá a zároveň poslední kapitola mé diplomové práce je zaměřena na vyhodnocení výsledků a formulaci závěrů.

1 DEFINICE ZÁKLADNÍCH POJMŮ VE ZVOLENÉ OBLASTI

V této kapitole budou vysvětleny základní pojmy týkající se RFM analýzy.

1.1 RFM analýza

Tato analýza je používána k přímé marketingové segmentaci. Základní otázkou v této analýze je, zda si zákazník zaslouží další kontaktování na základě jeho předchozího nákupního chování. Kontaktováním může být myšleno jak e-mailovou poštou, klasickou poštou, katalogem nebo telefonem. Jelikož ne všichni zákazníci objednávají stejně častěji a za stejné částky, neměli by pak být kontaktováni stejným způsobem a být vynakládány na ně stejné výdaje (Miglautsch, 2000, str.67).

Základními proměnnými této analýzy pak jsou (Rud, 2001, str. 160):

- R – Recency – Aktuálnost/Doba,
- F – Frequency – Frekvence/Četnost,
- M – Monetary – Peněžní hodnota.

R – Aktuálnost

Aktuálnost nebo jinak označováno *Doba* vyjadřuje počet uplynulých měsíců od posledního nákupu uskutečněným daným zákazníkem. Tato proměnná je ze všech tří proměnných nejvýznamnější v predikci budoucího nákupu. Přeci jen u zákazníka, který nakoupil u firmy v nedávné době je větší pravděpodobnost dalšího nákupu než u zákazníka nakupujícího naposledy například před 2 lety (Rud, 2001, str. 160).

F – Četnost

Četnost či *Frekvence* je udávána počtem nákupů ať již celkovým či za nějaké sledované období. Hned po *Aktuálnosti* se řadí jako druhá nejdůležitější proměnná (Rud, 2001, str. 160).

U zákazníků, kteří nakupují častěji a v menším intervalu je mnohem větší pravděpodobnost budoucího nákupu než u zákazníků, kteří uskutečnili jeden či dva nákupy za celou svou životnost u dané společnosti (Novo, 2000).

M – Peněžní hodnota

Peněžní hodnota udává celkovou utracenou sumu zákazníkem. Stejně jako u *Aktuálnosti* se buď jedná o celkovou hodnotu utracenou za celou existenci zákazníka, nebo o sumu utracenou za určité sledované období. S ohledem na předvídání budoucího nákupu tato

proměnná napovídá nejméně. Ačkoliv v kombinaci s dvěma předchozími proměnnými pomáhá získat lepší přehled o nákupním chování zákazníků (Rud, 2001, str. 161).

I přesto se ale musí vzít v potaz, že zákazník, který utratil celkově za všechny nákupy nejvíce peněz, bude s největší pravděpodobností nakupovat znovu. Neboť nejcennější zákazníci mají tendenci se stávat ještě cennějšími (Novo, 2000).

Jelikož je přímá marketingová segmentace věda, je důležité kvalifikovat chování zákazníků tak, aby krátkodobé i dlouhodobé efekty segmentace mohly být hodnoceny. Smyslem RFM analýzy je poskytnout jednoduchý rámec kvalifikování chování zákazníků. Jakmile jsou totiž zákazníci zařazeni do této analýzy, mohou být shlukováni do segmentů a jednotlivých sub segmentů. Toto rozdělení pak tvoří základ pro budoucnost v rozhodování o frekvenci a způsobu kontaktu s jednotlivými zákazníky (Miglautsch, 2000, str. 67).

1.1.1 Bodování RFM

Cílem bodování RFM analýzy je nastítnit budoucí nákupní chování zákazníků. Je podstatné převést toto chování každého zákazníka do čísla, které lze použít skrze časová období.

Velmi často bývá, že přímí marketéři používají statistické výběry zákazníků. Zde ale nastává problém. U prvotního tvoření tohoto segmentačního systému stanoví jako nejlepšího zákazníka, toho, kdo utratí nad určitou částku. Pokud je zákazník relativně nový, dojde bohužel ke zhoršení situace. Segment nejlepších zákazníků může prvotně obsahovat 20 % zákazníků, po jednom roce či dvou to už může být 30 až 40 %. Pokud ale tuto danou částku použije jako hranici pro nejlepší zákazníky znovu, přinese to horší výsledky.

Nejlepší zákazníci společnosti sice své chování nezmění, ale po nějakém čase stále více zákazníků bude opakovat hodnotu nákupu v dané výši. Následující 2 běžně používané metody jsou využívány k zabránění právě tomuto problému (Miglautsch, 2000, str. 67-68).

1.1.2 Kvantily zákazníků

Zákazníci v této metodě jsou rozděleni do 5 stejných kvantilů či skupin. Nejlepší zákazníci obdrží hodnocení 5 a nejhorší naopak 1. Pro aktuálnost (R) jsou zákazníci seřazeni podle měsíců od posledního nákupu, čím nižší počet, tím lepší hodnocení. Podle četnosti (F) jsou zákazníci seřazeni podle počtu nákupů, čím vyšší počet, tím lepší hodnocení. A pokud jde o peněžní hodnotu (M), jsou zákazníci seřazeni podle výše utracené částky, čím vyšší, tím opět lepší hodnocení.

Každým novým hodnocením se může vytvořit relativně nové schéma. Výhodná je zde tedy kvantifikace chování zákazníka, která lze promítnout do budoucnosti. Ti nejlepší zákazníci by ale vždy měli skončit v kategorii 5,5,5. Velmi klíčové je pak stanovit mezní hodnoty jednotlivých kategorií. Tyto mezní hodnoty se pak automaticky mění s každým hodnocením zákazníků.

Metoda kvantilů zákazníků má také výhodu v získávání stejného počtu zákazníků v každém segmentu. Existuje zde 5 stejných skupin generujících 125 stejně velkých segmentů (5x5x5). Dříve by přímí marketéři kontaktovali všechny zákazníky. Nyní s ohledem na jednotlivé skupiny, které mohou být například označeny 4,3,5 či 3,2,3, pochopíme, jak rozdílné tyto segmenty jsou a jaké zákazníky reprezentují a které tedy kontaktovat (Miglautsch, 2000, str. 68-69).

Tato metoda ale může také narazit na nějaká úskalí v pohledu na četnost (F). Vysoké procento zákazníků totiž objedná jen jednou. Toto procento se často pohybuje mezi 30 a 60 %. Pokud by tedy více než 20 % zákazníků mělo pouze jednu provedenou transakci, pak tato skupina bude mít číslo 1. Jelikož ale každá z 5 skupin může obsahovat pouze pětinu zákazníků, zbytek se přelije do skupiny s číslem 2 i přesto, že jejich chování je naprosto stejné. Při 40 % by pak skupiny 1 a 2 byly naprosto identické. Při zmiňované hranici 60 %, která není tak neobvyklá, by pak 3 z 5 skupin měly stejné chování. Vzhledem ke smyslu RFM analýzy by to poté nepřineslo moc uspokojivé výsledky. Dochází pak k tendenci seskupovat zákazníky, kteří mají výrazně odlišné chování (Miglautsch, 2000, str. 68-69).

1.1.3 Hodnocení kvantilů chování

Další metoda byla vyvinuta doktorem Johnem Wirthem. Tato metoda rovněž třídí zákazníky podle jejich chování, ale zde může být rozdílný počet zákazníků v jednotlivých kvantilech.

Bodování proměnné F pak tvoří 5 skupin, které jsou rozděleny následovně (WEI, 2010):

- 0 – 3 měsíců,
- 4 – 6 měsíců,
- 7 – 12 měsíců,
- 13 – 24 měsíců,
- 25 a více měsíců,

které jsou poté hodnoceny 5, 4, 3, 2 a 1. Tato metoda je také označována jako tvrdé hodnocení (WEI, 2010).

Pokud zákazník nakoupil jen jednou, je mu přiděleno skóre 1. Poté se zbylé hodnoty zprůměrují. Jakmile je celková četnost zákazníka nižší než průměr, je mu přiděleno skóre 2. U peněžní hodnoty je stále tvořeno 5 kvantilů, kdy každý má stejný počet nákupů (WEI, 2011).

Protože minulé chování je nejlepším předpokladem pro budoucí nákupního chování zákazníka, je aktuálnost obvykle považována za nejmocnější z těchto 3 proměnných. Mnoho přímých marketérů dělá rozhodnutí o kontaktování pouze na základě aktuálnosti. Berou se v úvahu nedávní zákazníci životaschopní jen po určitou dobu. Často jsou pravidelně kontaktováni v prvních 12 měsících a poté stále méně často do 36-48 měsíců. Po uplynutí této doby a stále neuskutečněného nákupu jsou považovány za mrtvé zákazníky (Miglautsch, 2000, str. 70).

1.1.4 Výhody RFM modelu

Existuje několik důvodů, proč je model RFM populární v segmentaci přímého marketingu po celá desetiletí. Výhody jsou následující (Wei, 2010):

- RFM model je nákladově efektivní při analýze zaměřené na chování zákazníků a pomáhá toto chování snadno kvalifikovat. Zákazníci a transakční data mohou být přitom ukládána v přístupných elektronických databázích,
- RFM je velmi cenný v predikci reakci zákazníka na marketing a může krátkodobě zvýšit zisky společnosti,
- Modelování je efektivní, neboť chování může být shrnuto pomocí velmi malého počtu proměnných,
- Proměnné RFM jsou shromažďovány prostřednictvím interní databáze obsahující specifické informace o zákaznících, týkající se historie transakcí a nemusí být tak získávány prostřednictvím souhrnných informací na úrovni demografické databáze. RFM je proto smysluplnější pro cílení na konkrétní zákazníky,
- RFM je dlouhodobě známá metoda pro měření síly vztahu se zákazníky, neboť může účinně identifikovat vysoce hodnotné zákazníky.

1.1.5 Nevýhody RFM modelu

Přestože je model RFM klíčovým nástrojem pro rozvoj firem v marketingové oblasti, má také i několik nevýhod (Wei, 2010):

- Jelikož je cílem RFM identifikovat cenné zákazníky, zaměřuje se tedy pouze na ty nejlepší. RFM poskytuje údaje jen o aktuálnosti, četnosti a hodnotě transakcí provedené zákazníkem. Ve skutečnosti ale 80 % nákupů uskuteční pouze 20 % zákazníků. Model tak ignoruje zákazníky, kteří spadají do kategorie 1,1,1, přičemž se jedná o nejpočetnější skupinu a tato skupina může mít tak největší nevyužitý potenciál,
- RFM model může používat pouze omezený počet výběrových proměnných. Nicméně ve většině případů mají největší vliv na odezvu od zákazníků na marketing společnosti,
- RFM se zaměřuje na současné zákazníky společnosti a nelze jej použít na vyhledávání nových,
- RFM podává jednu odpověď pro všechny zákazníky z databáze a předpokládá tedy homogenitu databáze. Což je často v rozporu se skutečností, kdy zákazníci mají značnou heterogenitu,
- RFM není považován za precizně kvalitativní model a důležitost každého jeho měření se liší skrze průmyslovými odvětvími.

1.1.6 Fáze zákazníků

Gary Seitz (2014) vysvětluje 4 různé fáze zákazníků dle RFM analýzy následovně.

Fáze 1 – Zkoušeči

Zákazníci této fáze jsou oceňujícím aktivem. Jsou novými zákazníky společnosti a je potřeba, aby nakoupili znovu. Tito zákazníci dobře reagují na marketing nejprodávanějších produktů obsahující i pokyny, jak tyto produkty používat. Důvěřují také referencím od spokojených zákazníků. Hlavním cílem je dostat je do fáze 2.

Fáze 2 – Kupující

Zákazníci druhé fáze jsou nejcennějším aktivem společnosti. Jedná se o věrné zákazníky, které si musíme udržet. Dobře reagují na častý marketing, který obsahuje nejnovější produkty. Jsou rádi odměňováni za jejich loajalitu a také rádi dotazováni na jejich názory či případné návrhy. Cílem je udržet si zákazníky právě v této fázi.

Fáze 3 – Vysoká hodnota, klesající potenciál

Tito zákazníci jsou odepisovaným aktivem. Potřebují být tedy znovu zaktivováni, neboť neodpovídají na marketing. Ať už jsou důvody k této změně jakékoliv, například přestěhování, úmrtí či přechod ke konkurenci, je potřeba tento důvod zjistit a pokud je to možné, odstranit jej.

Fáze 4 – Nízká hodnota, klesající potenciál

V této fázi už nejde o aktivum ale o závazek. Zákazníci této fáze jsou považovány za ztracené, neboť společnost už od nich nemá po dlouhou dobu žádnou odezvu na její marketing, a proto je potřeba je vyměnit. Pro společnost je proto v této fázi cílem přesměrovat své vynakládané zdroje od nich směrem k získání a udržení cennějších a aktivnějších zákazníků.

1.2 Určení RFM

Následující podkapitola se zabývá určováním RFM hodnoty. Jedná se o metody vážení a metody shlukování.

1.2.1 Metoda vážení

Stále častěji se stává, že si většina obchodníků vybírá právě skórování podle RFM analýzy pro své databáze zákazníků. Ostatní obchodníci však nemůžou spojit všechny 3 proměnné do jediné hodnoty. Přičemž hlavní výhodou jediné hodnoty je jednoduchá segmentace zákazníků. Vzorcem je pak pouhé sečtení těchto tří proměnných (Miglautsch, 2000, str. 71).

$$\text{RFM} = \text{R} + \text{F} + \text{M} \quad (1)$$

Tato metoda předpokládá sečtení aktuálnosti, frekvence a peněžní hodnoty. Jelikož je ale výše peněžní hodnoty ve srovnání s hodnotami aktuálnosti a frekvence vyšší, musí dojít k určení vah jednotlivých proměnných.

Alternativou by pak byl součet jednotlivých skóre každé proměnné. Při součtu těchto hodnot by pak nejlepší zákazníci měli skóre 15 (5 + 5 + 5) a nejhorší pak skóre 3 (1 + 1 + 1). Avšak mnoho zákazníků by mělo skóre 7 či 8 a bylo by obtížné je roztřídit či rozlišit.

Po desítkách let zkušeností přímý poštovní marketing naznačuje, že nejaktuálnější zákazníci jsou nejvíce hodnotní oproti zákazníkům, kteří ignorovali už několik e-mailů. Proto také spousta obchodníků násobí předchozí vzorec následovně:

$$\text{RFM} = R \times 3 + F \times 2 + M \times 1 \quad (2)$$

Nejlépeší zákazníci pak budou mít skóre 30 (5x3) + (5x2) + (5x1). Tento přístup dává nejen více síly nejnovějším zákazníkům, ale také zvyšuje sílu frekvence. Podstatou vážením frekvence, je skutečnost, že pokud by byly 2 zákazníci se stejnou aktuálností a peněžní hodnotou, ale jeden by nakoupil pouze jednou oproti druhému, který tak učinil několikrát. Při rozhodování, komu poslat katalog by bylo zřejmé poslat jej tomu, který nakoupil několikrát.

Dalším vylepšením je vytvoření složeného skóre, kdy místo násobení čísla 3, 2 a 1 se násobí 9,9; 6,6 a 3,3. Tím se získá větší rozsah skóre, a to mezi hodnotou 19,8 a 99. Stále by ale byla zachována 3násobnost R, ale vytvoří se tak téměř 100 bodová stupnice (Miglautsch, 2000, str. 71).

1.2.2 RFM a shlukování

Techniky shlukování spadají do skupiny nástrojů pro dolování dat. Cílem data miningu je objevit strukturu v datech jako celku. Neexistuje žádná cílová proměnná k předvídaní, takže se nerozlišuje mezi nezávislými a závislými proměnnými.

Shlukovací techniky jsou používány pro roztřídění prvků do klastrů neboli skupin, které splňují dvě hlavní kritéria (Aggelis, 2005, str. 2):

- Každý klaster je homogenní – prvky patřící do stejné skupiny či klastru jsou podobné,
- Každý klaster by měl být odlišný od ostatních skupin či klastrů – prvky patřící do jedné skupiny jsou odlišné od prvků z jiné skupiny.

V závislosti na technice shlukování mohou klastry být vyjádřeny různými způsoby (Aggelis, 2005, str. 2):

- Identifikované klastry mohou být výlučné – jakýkoliv prvek patří pouze jednomu klastru,

- Klasy se mohou překrývat – prvek může patřit do více klastrů,
- Klasy mohou být pravděpodobnostní – prvek patří do každého klasteru s určitou pravděpodobností,
- Klasy mohou mít hierarchickou strukturu – s hrubým rozdělením prvků na nejvyšší úroveň hierarchie, která je pak rozdělena do dílčích klastrů na nižších úrovních.

Metoda K-Means

Existuje celá řada metod na klastrování dat, jednou z nehierarchických je právě metoda K-Means, pomocí níž dochází k rozdělení dat na určitý počet klastrů k . Při tomto rozdělení jsou pak data v daném klasteru vysoce podobná a zároveň jednotlivé klasy a data v nich vzájemně rozdílná. U této metody lze rovněž pracovat i s jinými atributy než pouze s R, F, M.

Metoda K-Means často používá euklidovský vzorec vzdálenosti k určení podobnosti data v klasteru (Gustriansyah a spol., 2020, str. 471).

K-Means je jedním z nejpoužívanějších metod shlukování, které jsou založené na participaci. Jedná se o velmi jednoduchý a rychlý algoritmus pro výběr hlavních klastrů (Bholowalia a Kumar, 2014, str. 18).

Jednotlivé kroky K-Means algoritmu jsou pak následující (Aggelis, 2005, str.2):

- 1) Určení center klastrů (centroidy),
- 2) Náhodné umístění centroidů v prostoru dat,
- 3) Přiřazení dat k nejbližším centroidům, čímž vznikne k shluků,
- 4) Přepočítání centroidů shluků tak, aby šlo o těžiště objektů daného shluku,
- 5) Kroky 3 a 4 se opakují, dokud nedojde k ustálení.

Jedním z hlavních problémů metody K-Means je, jak určit optimální počet klastrů k . Existuje několik způsobů, které lze použít k odhadu optimálního počtu klastrů k . Jako je například Elbow Method, Silhouette Index, Calinski-Harabaszův Index nebo Ball-Hall Index (Gustriansyah a spol., 2020, str. 471).

1.3 Hodnota pro zákazníka

V posledních letech jsou zákazníci mnohem vzdělanější a informovanější, než jak tomu bylo dříve. Zákazníci mají totiž různé nástroje k ověření tvrzení společnosti a najít tak případných lepších alternativ u konkurenčních společností. Hledají maximální vnímanou hodnotu, kterou mohou získat při přiměřených nákladech, dosavadních znalostech a příjmu. Konečná rozhodnutí pak dělají na základě toho, která nabídka jim tuto nejvyšší hodnotu poskytne (Kotler a Keller, 2007, str. 179).

Hodnota pro zákazníka pak představuje očekávání zákazníka, že jim celková nabídka poskytne něco navíc a bude pro něj hodnotná (Lošťáková, 2017, str. 39).

S touto tematikou pak přicházejí následující tři pojmy, a to Hodnota vnímaná zákazníkem, Celková hodnota pro zákazníka a Celkové náklady pro zákazníka.

Hodnota vnímaná zákazníkem neboli CPV (customer perceived value) se rozumí jako rozdíl mezi všemi výhodami plynoucí z nabídky a náklady na tuto nabídku s ohledem na všechny nabízené varianty.

Celková hodnota pro zákazníka pak už představuje peněžní vyjádření daných výhod, ať už funkčních, psychických či ekonomických zvolené nabídky.

Celkové náklady pro zákazníka představují veškeré náklady, které musel zákazník vynaložit při shromáždění, hodnocení, využívání a zbavování se dané nabídky. Tyto náklady pak mohou být finančního, časového, energetického ale i psychického charakteru (Kotler a Keller, 2007, str. 179).

Proto, aby společnost vytvářela dobré a dlouhotrvající vztahy se zákazníky, je kritické pochopit, co je hlavním cílem zákaznickova nákupu. Například jaké vlastnosti, užítky a atributy zákazník vyhledává a preferuje. Nabídkou by pak měl být souhrnný komplex užitek, které zajistí zákazníkovi co nejvyšší hodnotu (Lošťáková, 2017, str. 39).

Wirtz a spol. (2018, str. 88-107) zpracovali a vysvětlili základní oblasti, kde může společnost zvýšit hodnoty pro zákazníka. Tyto oblasti nazývají jako 8 klíčových bodů přídavných služeb, které zvedají hodnotu spolu se základním zakupovaným produktem či službou. Zmiňované oblasti jsou následující:

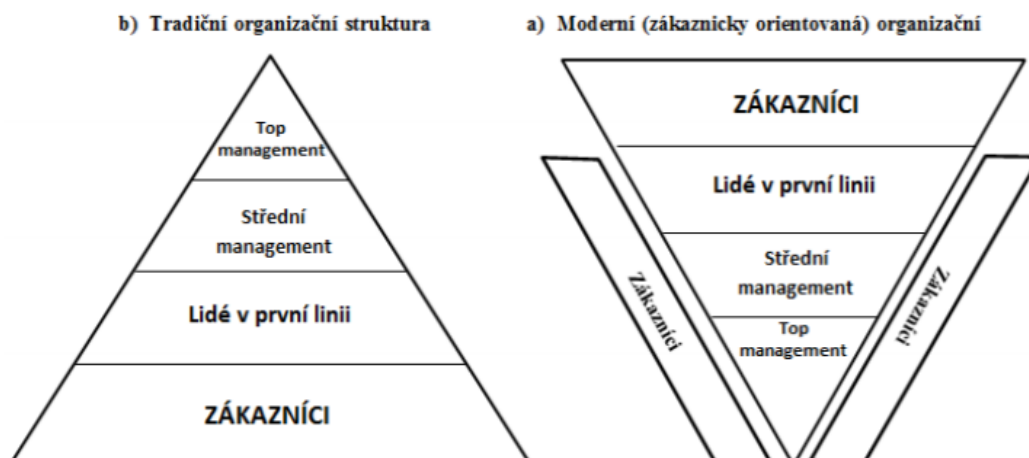
- **Poskytování informací** – jen tehdy, pokud společnost poskytne zákazníkovi úplné a relevantní informace o produktu či službě, může dosáhnout zákazník nejvyšší hodnoty. To obzvláště platí, pokud zákazník nakupuje produkt poprvé,

- **Poradenské služby** – ačkoliv už od společnosti zákazník dostal prvotní informace, je důležité zůstat se zákazníkem v kontaktu. Pomoci mu s jakýkoliv požadavkem či problémem a poskytnout případné konzultace,
- **Zpracovávání objednávek** – velkou roli hraje také objednávkový systém společnosti. Jestliže se zákazník rozhodne o uskutečnění nákupu, měl by být tento systém rychlý, bezproblémový a spolehlivý,
- **Komfort pro zákazníky, pohoštění a pohodlí** – vysvětluje podstatu převzetí zákazníka do své péče a nabídnout mu prostředí, kde by se cítil jako vítaný host,
- **Zabezpečení bezpečnosti zákazníků** – obsahuje jasná pravidla, ať už se týkají odstoupení od smlouvy, záruky, vrácení produktů či pozáruční servis,
- **Nadstandardy** – zaměřuje se na všechny formy a způsoby zvláštních či mimořádných nabídek a péči o zákazníky,
- **Bezproblémová fakturace** – týká se opatření k zajištění poslání úplných a bezchybných faktur pro úplnou spokojenost zákazníka,
- **Jednoduché placení** – zaměřené na automatizaci plateb za produkty či služby, tak aby to co nejméně zatěžovalo nakupujícího zákazníka.

Přístup k zákazníkům

Novodobé úspěšné společnosti pokládají tradiční pyramidovou strukturu podniku za zastaralou. Tato struktura pokládala na vrchol pyramidy top management, poté střední management v závěsu s lidmi v přední linii a se zákazníky úplně vespod. Současní manažeři, ale pochopili, že opravdovým centrem zisku jsou zákazníci, a proto tuto pyramidu kompletně otočili.

Na samotný vrchol tedy postavili zákazníky, které následují lidé v přední linii, kteří dbají o jejich spokojenost a potřeby. Střední management zde slouží jako podpora pro lidi v přední linii. Na závěr je pak vespod top management, jejichž úkolem je získávat a podporovat schopné manažery střední úrovně. Mnoho nových společností je již zakládáno s touto strukturou viz. obrázek 1, kde se staly zájmy a potřeby zákazníků jejich strategií a zároveň i konkurenční výhodou (Kotler a Keller, 2007, str. 178-179).



Obrázek 1 Tradiční a moderní zákaznický orientovaná struktura

Zdroj: Kotler a Keller, 2007, str. 178

1.4 Segmenty zákazníků

Přístupů k segmentaci zákazníků v RFM analýze je spousta. Jedním z nich je například následující, který člení zákazníky na osm segmentů (Birant, 2011, str. 102-103):

1. Nejlepší zákazníci,
2. Hodnotní zákazníci,
3. Nakupující zákazníci,
4. Nakupující poprvé,
5. Chrlící zákazníci,
6. Častí zákazníci,
7. Utrácející zákazníci,
8. Nejistí zákazníci.

Segment **nejlepších zákazníků** je ten nejcennější, skládá se ze zákazníků, kteří nedávno provedli pravidelné nákupy, mají vyšší průměrnou četnost nákupů i vyšší průměr peněžní hodnoty.

Hodnotní zákazníci jsou hned v závěsu za nejlepšími. I u nich je průměrná aktuálnost, četnost a peněžní hodnota vyšší než celková průměrná hodnota.

U **nakupujících zákazníků** je opět průměrná hodnota aktuálnosti a četnosti vyšší než celková průměrná hodnota. Průměrná peněžní hodnota je zde ale nižší.

Segment **nakupujících poprvé** zahrnuje zákazníky, kteří nedávno poprvé navštívili společnost, a tedy mají vyšší průměrnou aktuálnost ale nižší průměrnou četnost a peněžní hodnotu (Birant, 2011, str. 102-103).

Chrlící zákazníci provádějí vysoký počet nákupů s vysokými peněžními hodnotami ale ne na dlouhou dobu. U těchto zákazníků se mohlo něco pokazit, a proto je nutné je kontaktovat a naplánovat reaktivační program, respektive návrh propagace.

Častí zákazníci jsou charakterističtí vyšší průměrnou četností, ale nižší aktuálností a peněžní hodnotou.

Utrácející zákazníci mají vyšší průměrnou peněžní hodnotu, ale jejich aktuálnost a četnost je nižší než celkový průměr.

Posledním segmentem jsou **nejistí zákazníci**, který je považován za nejméně cenným pro společnost. Zahrnuje zákazníky, kteří mají skóre RFM 1,1,1; 1,1,2 či 1,2,1 a zde je obecně nejmenší pravděpodobnost, že by nakoupili znovu.

Segmenty, které mají alespoň dvě hodnoty vyšší než celkový průměr těchto hodnot, lze vybrat jako cílové. Zákazníci v těchto segmentech se stávají kandidáty na vedení vhodné marketingové strategie, které přitahují největší pozornost (Birant, 2011, str. 102-103).

Jiný přístup k rozdělení zákazníků pak popisuje Štráfelda (2021), který jmenuje těchto následujících 5 segmentů:

- Noví zákazníci,
- Zlatí zákazníci,
- Loajální zákazníci,
- Upadající zákazníci,
- Ztracení zákazníci.

Noví zákazníci, jak z názvu vyplývá nakoupili u společnosti v nedávné době a pouze jednou. Cílem společnosti tedy bude přimět je nakoupit znovu. Těmto zákazníkům také může pomoci společnost zorientovat se v jejich nabídce formou newsletteru či mailingu nastaveným k pravidelnému zasílání po uplynutí určité doby od nákupu. Dále tito zákazníci mohou ocenit návody a tipy k zakoupeným výrobkům a jejich údržbě a možnosti servisu.

Zlatými zákazníky pak jsou zákazníci, kteří uskutečňují objednávky za vysoké částky, velmi často, pravidelně a v nedávné době. Jedná se o nejlepší zákazníky společnosti.

Společnost by se měla snažit s nimi navázat osobitější přístup například přidělením osobního poradce, vytvoření speciálního klubu či nabídek šitých na míru. Tyto zákazníky je také vhodné využít k šíření pozitivních recenzí, a tak dát o své společnosti více vědět.

Loajální zákazníci provádí objednávky často, pravidelně ale ne v takové výši jako zákazníci zlatí. Případným začleněním jich do věrnostních programů mohou společnosti ještě více navýšit jejich loajalitu vůči nim.

Upadajícími zákazníky rozumíme zákazníky, kteří u společnosti pravidelně nakupovali ale v poslední době o nich společnost neslyšela. Cílem je pak tedy takového zákazníka přimět znovu nakoupit. Jestliže se toto společnosti nepodaří o zákazníka tak může definitivně přijít. Jednou z možností, jak toho dosáhnout je nabídnout výrazné slevy či časově omezené nabídky.

Ztracení zákazníci už po dlouhou dobu žádnou objednávku neuskutečnili. Jestliže dříve patřili ke zlatým zákazníkům, může se společnost pokusit o jejich oživení, v opačném případě je lepší se tohoto segmentu vzdát a investovat do lukrativnějších segmentů (Štráfelda, 2021).

2 FORMULACE PROBLÉMU

Tato kapitola obsahuje představení společnosti, na základě jejíž poskytnutých dat bude provedena RFM analýza.

Pro tuto diplomovou práci byla poskytnuta reálná data objednávek za období od začátku roku 2018 do konce října roku 2020.

Společnost poskytující data působí v odvětví zabývající se výrobou a rozvozem ovocných květin, klasických květin a dortů. Společnost byla založena roku 2012 a nyní zaměstnává 15 stálých zaměstnanců. Společnost má na území České republiky dvě pobočky, a to v Praze a v Brně. Nabízí rozvoz svých výrobků v rámci celé republiky, kdy objednávka může být uskutečněna odkudkoliv na světě.

V současnosti společnost segmentuje své zákazníky na pouze dva segmenty, a to na nové a vracející se zákazníky. Do segmentu „vracející se“ zahrnuje jak zákazníky, kteří objednávku už jednou uskutečnili, tak i zákazníky, kteří navštívili její webové stránky. Na tento segment pak používá remarketing.

Remarketingem se rozumí způsob, jímž se společnosti snaží přilákat zpátky zákazníky na své stránky a uskutečnit nákup, který při prvotní návštěvě nedokončili (Computer Press, 2014, str. 61).

Prvotní data pak byla poskytnuta prostřednictvím souboru v MS Excel. Tato data se skládala ze čtyř atributů: ID objednávky, Jméno zákazníka, Hodnota objednávky a Datum objednání. ID objednávky je jedinečným číslem pro každou objednávku. Jméno zákazníka je tvořeno křestním jménem a příjmením zákazníka. Hodnota objednávky je číselná hodnota nabývajících hodnot od 0 Kč po 45 049 Kč. Datum objednání je dáno datem a časem každé objednávky. K dispozici bylo celkem 49 603 objednávek. Následující tabulka 1 znázorňuje datový slovník prvotních atributů.

Tabulka 1 Datový slovník prvotních dat

Atribut	Popis	Formát	Rozsah
ID objednávky	Jedinečné číslo pro každou objednávku	Číslo	
Jméno zákazníka	Křestní jméno a příjmení zákazníka	Text	
Hodnota objednávky	Cena objednávky vyjádřena v českých korunách	Číslo	0 – 45 049 Kč
Datum objednání	Datum a čas objednávky	Datum a čas	1.1.2018 02:25 – 31.10.2020 14:36

Zdroj: vlastní zpracování

Na obrázku 2 je znázorněna ukázka prvotních poskytnutých dat.

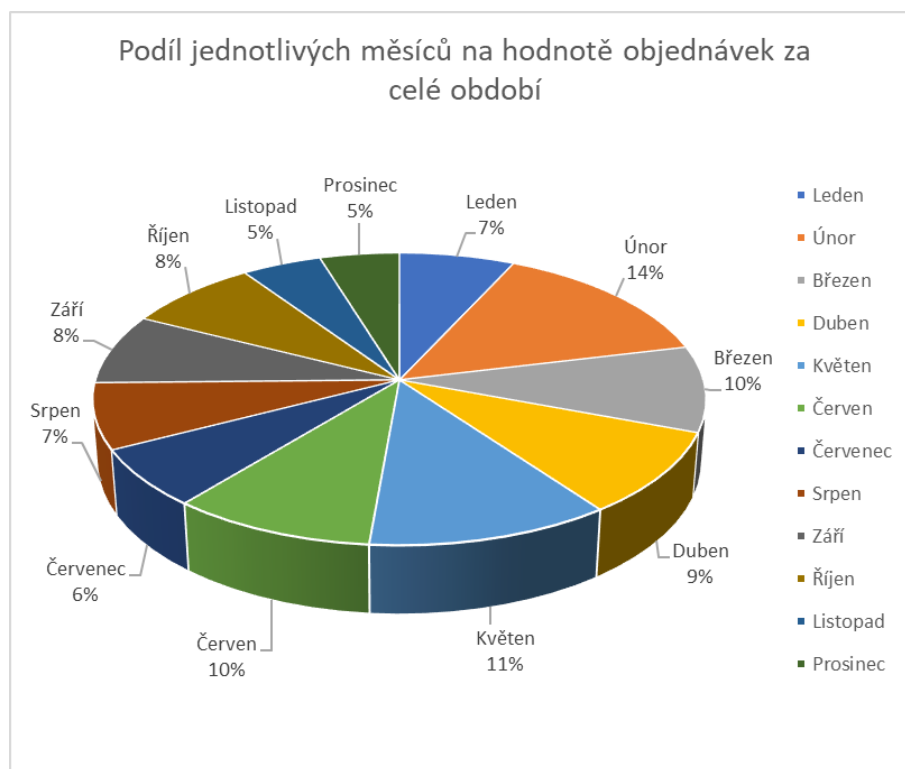
ID objednávky	Jméno zákazníka	Hodnota objednávky	Datum objednání
62072		849,00 Kč	01.01.2018 2:25
62073		427,00 Kč	01.01.2018 4:19
62074		1 168,00 Kč	01.01.2018 11:02
62075		549,00 Kč	01.01.2018 12:11
62076		4 147,00 Kč	01.01.2018 12:20
62077		1 618,00 Kč	01.01.2018 17:44
62078		1 198,00 Kč	01.01.2018 19:51
62079		899,00 Kč	01.01.2018 20:05
62080		1 068,00 Kč	01.01.2018 20:08
62081		698,00 Kč	01.01.2018 20:17
62082		768,00 Kč	01.01.2018 21:19
62083		578,20 Kč	01.01.2018 21:26
62084		1 357,00 Kč	01.01.2018 21:46
62085		998,00 Kč	01.01.2018 21:57
62086		1 916,00 Kč	01.01.2018 21:59
62087		1 068,00 Kč	01.01.2018 21:59
62089		1 367,00 Kč	01.01.2018 22:18
62090		1 238,00 Kč	02.01.2018 7:52

Obrázek 2 Ukázka prvotních poskytnutých dat

Zdroj: Interní zdroje, 2020

Vizualizace dostupných dat

Jak již bylo zmíněno, výše poskytnuté informace o objednávkách se týkají období od začátku roku 2018 až do konce října roku 2020. Na následujícím obrázku 3 jsou znázorněny podíly jednotlivých měsíců na celkové hodnotě objednávek za toto období.

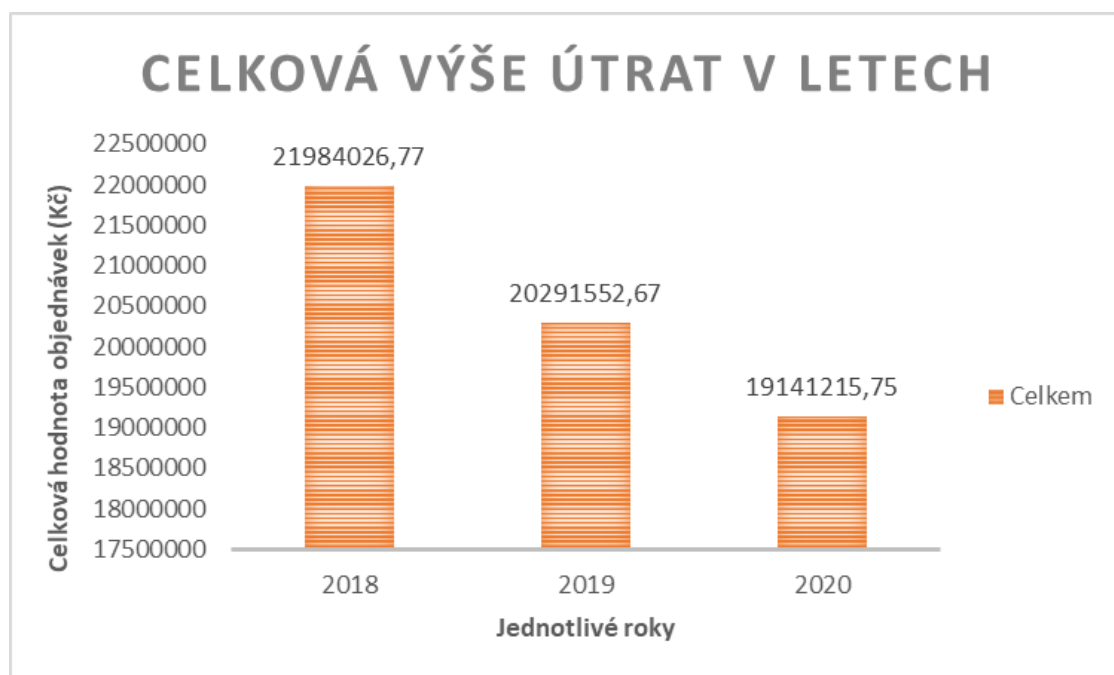


Obrázek 3 Podíl měsíců na celkové hodnotě za dané období

Zdroj: vlastní zpracování

Na základě jednotlivých procentních zastoupení měsíců je patrné, že měsíc nemá až takový vliv na výši hodnot objednávek. Největší procentní zastoupení má ale měsíc květen s 11 %, naopak nejméně procentního zastoupení pak měly měsíce listopad a prosinec.

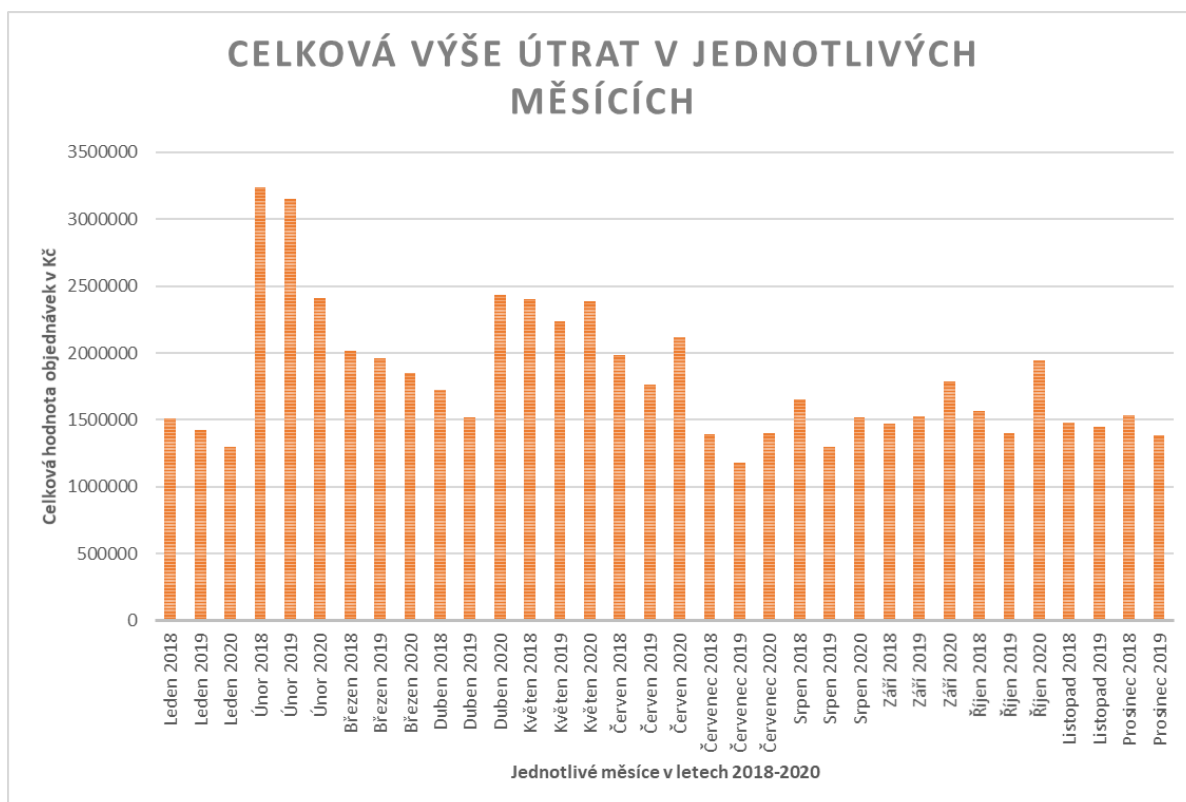
Na následujícím obrázku 4 je znázorněna celková výše útrat za jednotlivé roky. Je zřejmé, že společnost v roce 2018 zaznamenala nejvyšší tržby za poslední 3 roky, a to téměř 22 milionů korun českých. Sice nejsou známy objednávky za poslední 2 měsíce v roce 2020, ale z výše zmíněného grafu je patrné, že právě tyto měsíce mají nejnižší celkovou hodnotu objednávek, a proto by ani celkovou výši za tento rok dostatečně nezvýšily.



Obrázek 4 Celková výše útrat v letech 2018-2020

Zdroj: vlastní zpracování

Na obrázku 5 jsou pak znázorněny útraty za jednotlivé měsíce za dané období. Z grafu je zřejmé, že nejvyšší tržby společnost zaznamenala v únoru roku 2018, které činily přes 3 200 000 Kč. Přes hodnotu 3 000 000 Kč se opět dostala v měsíci únor hned následující rok 2019. Naopak nejnižší tržby zaznamenala společnost v červenci roku 2019, které činily bez mála 1 200 000 Kč. Při pohledu na graf je dále patrné, že tržby v zimních měsících od ledna do března mají klesající vývoj, naopak zbylé měsíce od dubna po říjen mají v roce 2020 rostoucí trend oproti předchozím rokům.



Obrázek 5 Celková výše útrat v jednotlivých měsících

Zdroj: vlastní zpracování

Cílem diplomové práce je pak využití RFM analýzy na poskytnutých datech a rozdělení zákazníků na více segmentů, než je tomu nyní.

Pro tuto diplomovou práci jsem se rozhodla vytvořit 3 různé analýzy. V prvním případě budou ponechány hraniční hodnoty určené softwarem (přístup 1, RFM), v druhém případě bude využita možnost nastavení vlastních hraničních hodnot (přístup 2, RFM) a ve třetím případě bude využito shlukové analýzy (přístup 3, K-Means).

3 ZPRACOVÁNÍ DAT POMOCÍ RFM ANALÝZY

Následující kapitola obsahuje představení softwaru, pomocí kterého budou dostupná data zpracována. Dále se zaměřuje na přípravu dat a samotné provedení analýzy na dostupných datech.

3.1 IBM SPSS Modeler

IBM SPSS Modeler je předním řešením pro vizualizaci dat a strojového učení. Pomáhá podnikům přeměnit čas v hodnotu a dosáhnout tak požadovaných výsledků zrychlováním provozních úkonů pro pracovníky s daty. Přední celosvětové organizace spoléhají na tento software při přípravě a zjišťování dat, predikční analytice, správě a implementování modelů a při zpeněžování datových aktiv. IBM SPSS Modeler umožňuje organizacím využívat datových aktiv a moderních aplikací s kompletními algoritmy a modely k okamžitému použití (IBM, 2021).

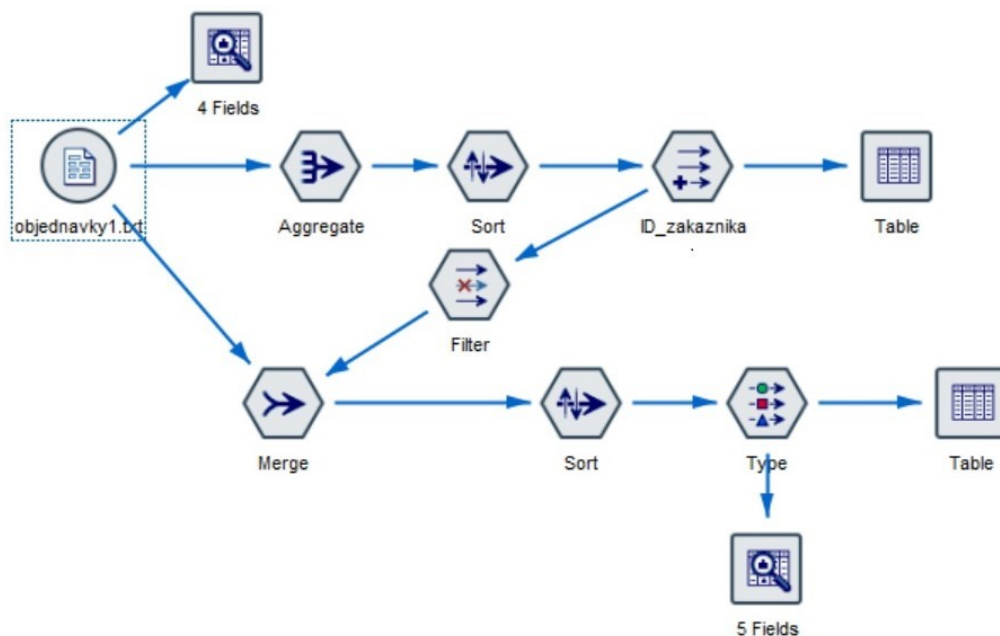
Mezi klíčové vlastnosti softwaru pak patří (IBM, 2021):

- Podpora mnoha zdrojů dat – dokáže číst data jak z plochých souborů, tabulek, velkých souborů tak i z relačních databází,
- Automatická příprava dat – automaticky transformuje data do nejlepšího formátu pro co nejpřesnější prediktivní modelaci. Během pár kroků tak vytvoří analýzu dat, identifikaci oprav, vyčistí pole a odvodí nové atributy,
- Výkonný grafický modul – obsahuje funkci pro doporučení grafu z desítek možností, který se nejvíce hodí pro daná data pro rychlé a efektivní sdílení poznatků a postřehů,
- Streamy vizuální analýzy – poskytuje intuitivní grafické rozhraní, které pomáhá vizualizovat každý krok v procesu dolování dat jako součástí streamu, takzvaného proudu,
- Automatizované modelování – může testovat více metod modelování, porovnávat výsledky a vybrat vhodný model v jediném kroku. To umožňuje rychle vybrat nejvýkonnější algoritmus na základě výkonu modelu,
- Velká škála algoritmických metod – nabízí několik technik strojového učení, včetně klasifikačních, segmentačních a asociačních algoritmů,
- Textová analytika – zachycuje klíčové koncepty, témata a trendy pomocí analýzy nestrukturovaných dat. Přináší pak tedy možnost odhalit cenné poznatky o obsahu blogu, zpětné vazbě zákazníků, emailů či o komentářích na sociálních sítích,

- Geoprostorová analytika – pomocí IBM SPSS Modeleru lze také prozkoumat geografická data a spojením těchto informací s aktuálními a historickými údaji tak generovat lepší přehlednost a přesnost predikce,
- Metody a algoritmy strojového učení – podporuje rozhodovací stromy, neuronové sítě a regresní modely. Lze zde využít přenosné funkce s prediktory a detekcí odlehlých hodnot dále také těžit z hierarchických modelů, kauzálního modelování a využívání časových řad.

3.2 Příprava dat

Před zahájením samotné analýzy muselo nejdříve dojít k úpravě prvotních dat. Základními atributy potřebnými pro analýzu jsou ID zákazníka, datum nákupu a hodnota nákupu. Poslední 2 atributy se v prvotních poskytnutých datech vyskytují, ID zákazníka ale nikoliv. Bylo potřeba tedy data pročistit a přiřadit ke každé objednávce ID zákazníka. Tato úprava byla provedena pomocí výše zmiňovaného softwaru následujícím streamem neboli proudem znázorněným na obrázku 6.



Obrázek 6 Stream pro přiřazení ID zákazníka

Zdroj: vlastní zpracování

Po spuštění vytvořeného streamu vznikl soubor, kde bylo přiřazeno ID zákazníka viz obrázek 7. Tento soubor ale ještě nemohl být použit k dalšímu zpracování. Systém přiřazující

ID totiž nevzal v potaz, když zákazník zadal jednou jméno s diakritikou a jindy naopak bez, či rozdíl mezi malými písmeny a velkými písmeny. Bylo tak přiděleno více ID ke stejnému zákazníkovi, proto musela být přidělená ID manuálně pročištěna. Po tomto kroku již mohl být soubor použit k dalšímu zpracování a pro vytvoření analýzy jsou již nyní k dispozici všechny 3 potřebné atributy. Z původních 49 603 objednávek a 32 636 různých ID bylo ve výsledku po pročištění k dispozici 49 437 objednávek od 30 433 různých zákazníků.

Table (5 fields, 49 603 records) #2

File Edit Generate

Table Annotations

	Jméno zákazníka	ID_zakaznika	ID objednávky	Hodnota objednávky	Datum objednání
1		10085	100325	2997.000	2020-01-16 07:28:00
2		14283	90666	1548.000	2019-05-22 08:55:00
3		13105	93314	1448.000	2019-07-20 11:14:00
4		15231	88441	998.000	2019-04-12 07:44:00
5		7930	105314	1397.000	2020-04-05 12:58:00
6		29844	108436	748.000	2020-05-15 14:18:00
7		23364	74610	238.000	2018-07-18 07:34:00
8		11439	97147	1598.000	2019-10-23 10:16:00
9		30797	107665	64.000	2020-05-06 02:10:00
10		20587	114959	1247.000	2020-09-22 11:09:00
11		22144	76451	1649.000	2018-08-29 11:11:00
12		25895	71009	1048.000	2018-05-12 15:32:00
13		9788	101041	1596.000	2020-02-02 06:44:00
14		26479	70153	848.000	2018-05-02 12:25:00
15		9430	101820	848.000	2020-02-11 10:48:00
16		15981	86701	599.000	2019-03-08 15:10:00
17		27509	68685	1499.000	2018-04-05 10:23:00
18		16567	85444	699.000	2019-02-19 12:16:00
19		28070	109953	1148.000	2020-06-08 20:39:00
20		19432	80610	1888.000	2018-12-04 12:16:00
21		19653	80241	748.000	2018-11-26 14:50:00
22		25759	71203	1486.000	2018-05-14 16:02:00
23		22040	76613	1498.000	2018-09-01 12:00:00
24		26129	111490	798.000	2020-07-07 08:53:00

Obrázek 7 Ukázka dat s přidělením ID zákazníka

Zdroj: vlastní zpracování

V rámci softwaru IBM SPSS Modeler lze také získat datový audit. Datový audit znázorňuje vybrané základní charakteristiky jednotlivých atributů, jako je například minimální a maximální hodnota, průměrná hodnota, směrodatná odchylka ale i koeficient šikmosti. Dále je zde vizualizace dat. V posledních sloupcích jsou pak udána celková množství dat a počet jedinečných hodnot. Tento datový audit je znázorněn na následujícím obrázku 8.

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
Jméno zákazníka		Nominal	--	--	--	--	--	32636	49603
ID_zakaznika		Nominal	1	32636	--	--	--	32636	49603
ID objednávky		Nominal	62072	117116	--	--	--	49603	49603
Hodnota objednáv...		Continuous	0.000	45049.000	1225.002	735.414	10.926	--	49603
Datum objednání		Continuous	2018-01-0...	2020-10-31 14:36:00	--	--	--	--	49603

Obrázek 8 Prvotní datový audit

Zdroj: vlastní zpracování

3.3 RFM v prostředí IBM SPSS Modeler

V IBM SPSS Modeleru se nacházejí dva základní uzly týkající se RFM analýzy, a to RFM Aggregate a RFM Analysis. Oba uzly jsou nastaveny tak, aby používaly nezávislé binování neboli seskupování hodnot do menších intervalů.

Uzel RFM Aggregate

Tento uzel umožňuje vytvoření nových atributů R, F, M vyplývajících z historických transakčních dat zákazníků. Tyto atributy pak uvádí, kdy byli jednotliví zákazníci naposledy u společnosti aktivní (R), kolik transakcí provedli (F) a celkovou hodnotu těchto transakcí (M) (více viz. IBM Documentation, 2021a).

Uzel RFM Analysis

Po získání jednotlivých hodnot R, F a M ke každému zákazníkovi může být využit uzel RFM Analysis pro další zpracování. Do tohoto uzlu se pak zadávají hodnoty R, F a M.

Následuje nadefinování počtu skupin (výchozích je 5 skupin), které mají být pro R, F a M vytvořeny. Po nastavení počtu skupin se nastavují váhy. Ve výchozím nastavení je nejvyšší důležitost přiřazena údajům o aktuálnosti s váhou 100, následuje četnost s váhou 10 a na závěr peněžní hodnota s váhou 1. V případě potřeby lze ale jednotlivé váhy upravit a stanovit tak nejvyšší důležitost hodnotě jiné.

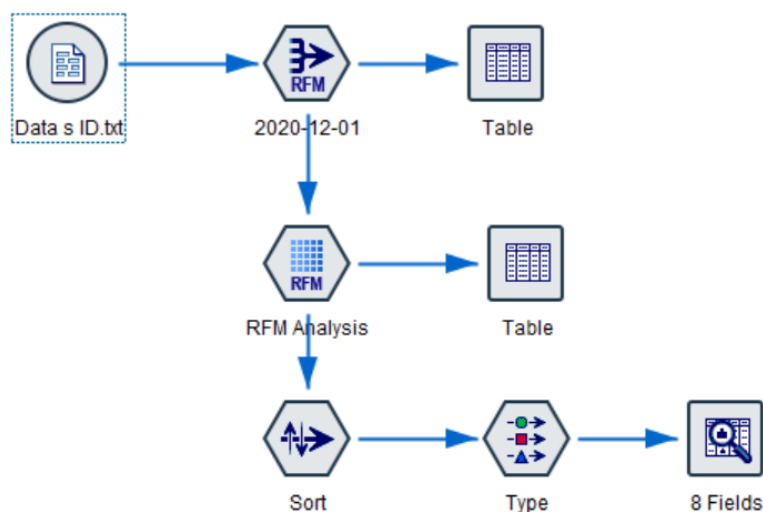
Po nastavení vah pro jednotlivé hodnoty dochází k nastavení hraničních hodnot, zde software nabízí dvě možnosti. V prvním případě jsou hraniční hodnoty stanoveny automaticky systémem a nelze je měnit. Druhou možností je schopnost nastavení vlastních hraničních hodnot (IBM Documentation, 2021b).

Po spuštění uzlu RFM Analysis dojde k vytvoření skóre aktuálnosti, četnosti, peněžní hodnoty a výsledného RFM skóre, které představuje vážený součet skóre aktuálnosti, četnosti a peněžní hodnoty.

V případě existence odlehlých hodnot lze nastavit jejich zařazení do koncových skupin. Data pak ležící pod nejspodnější skupinou jsou přidána do spodní skupiny a data ležící nad nejvyšší skupinou jsou přidána do skupiny nejvyšší. Tato možnost je možná ale jen u zvolení varianty vlastního nastavení hraničních hodnot (IBM Documentation, 2021b).

3.4 Zpracování dat

Pročištěný soubor dat s přiděleným ID bude nyní sloužit jako vstupní datový soubor pro stream obsahující samotnou RFM analýzu. Stream znázorněný na následujícím obrázku 9 přidělí jednotlivým zákazníkům hodnoty R – Aktuálnost, F – Četnost a M – Peněžní hodnota.



Obrázek 9 Stream RFM analýzy

Zdroj: vlastní zpracování

Jednotlivé hodnoty R, F a M jsou znázorněny na obrázku 10 (příklad 20 zákazníků a hodnot R, F, M). Analýza byla provedena k datu 1. prosince 2020. Z celkových 49 437 objednávek se jedná o 30 433 různých zákazníků. V průměru tedy na každého zákazníka připadá 1,62 objednávky.

	ID zákazníka	Recency	Frequency	Monetary
1	16091	636	1	748.000
2	25532	926	1	848.000
3	19253	722	1	1278.100
4	26323	940	1	1198.000
5	18865	708	1	748.000
6	4816	657	2	577.000
7	32635	1065	1	427.000
8	5442	200	2	3443.000
9	24063	117	1	798.000
10	5846	135	2	1365.000
11	17614	661	1	1498.000
12	1736	503	4	10739.0...
13	29538	196	1	698.000
14	30436	1024	1	586.000
15	2152	35	3	2679.070
16	9394	294	1	1597.000
17	24309	896	1	1099.000
18	9057	291	1	2047.000
19	14463	567	1	1847.000
20	14548	570	1	1248.000

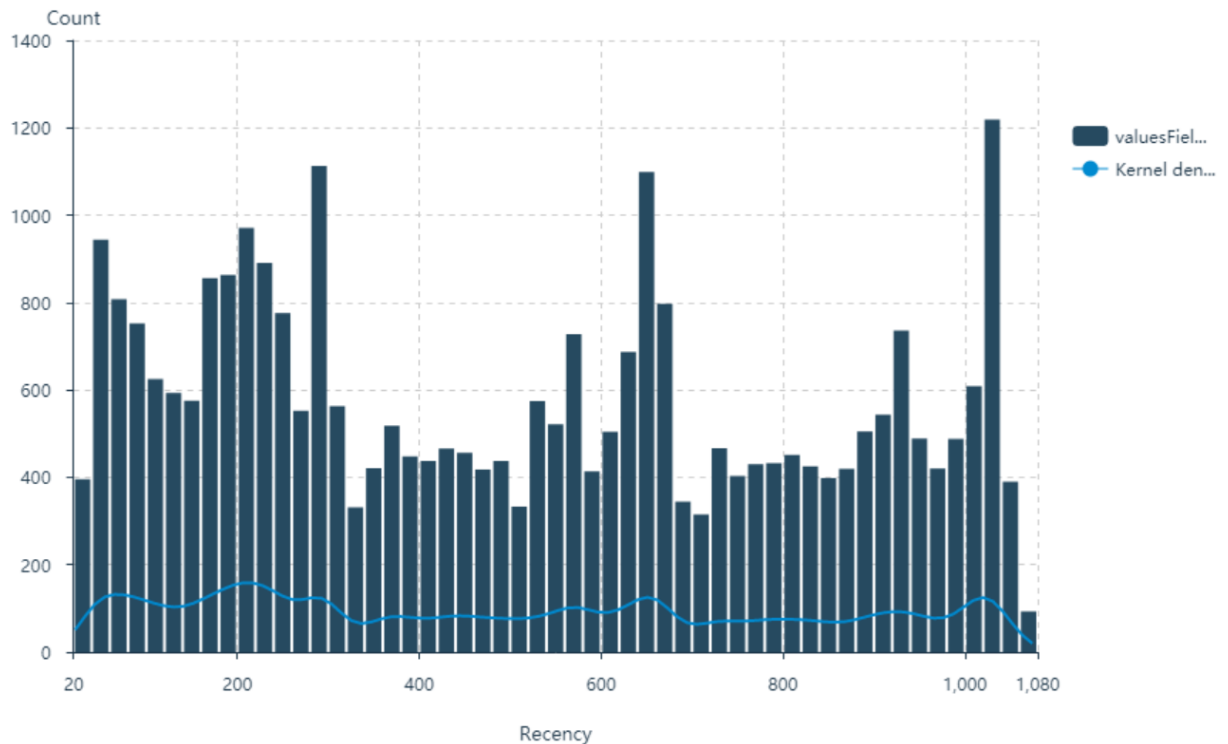
Obrázek 10 Hodnoty R, F a M

Zdroj: vlastní zpracování

Nyní budou znázorněny a okomentovány histogramy a krabicové grafy neboli *boxplots*, jednotlivých hodnot R, F a M, které lze v prostředí IBM SPSS Modeler vytvořit.

Histogram Recency

Na následujícím obrázku 11 je znázorněn histogram hodnoty R (Aktuálnost). Na ose x jsou uvedeny počty dnů od posledního nákupu a na ose y pak počet jednotlivých zákazníků, kteří této hodnoty dosahují. Jednotlivé sloupce histogramu jsou ve 20denních intervalech. Minimální zaznamenanou hodnotou neboli nejnižší počet dnů od poslední objednávky je 31 dní. Tato hodnota se nachází v prvním intervalu představující 20. – 40. den s četností 396 zákazníků, kteří tak učinili. Naopak maximální zaznamenaná hodnota a tím pádem nejvyšší počet dnů od poslední objednávky je 1065 dní znázorněná v intervalu 1 060. – 1 080. den představující 93 různých zákazníků, kteří takto naposledy objednali.

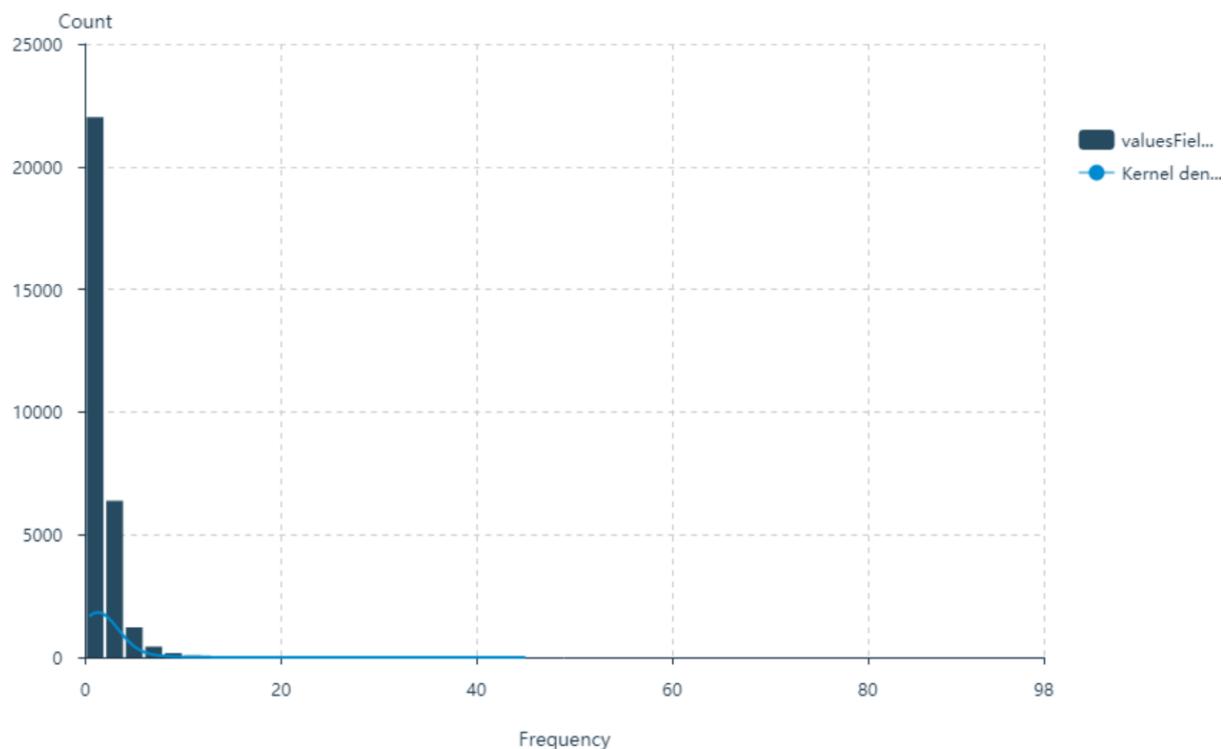


Obrázek 11 Histogram hodnoty R

Zdroj: vlastní zpracování

Histogram Frequency

Následující histogram na obrázku 12 znázorňuje hodnoty F (Četnost), kdy na ose x je znázorněn počet uskutečněných objednávek a na ose y počet zákazníků uskutečňující daný počet. Jednotlivé intervaly histogramu jsou po 2 objednávkách. Z histogramu je zřetelné, že nejvíce zákazníků uskutečnilo 0 – 2 objednávky a to přesně 22 009 zákazníků, další výraznější hodnotou je 6 382 zákazníků, kteří uskutečnili 2 – 4 objednávky. S přibývajícím počtem objednávek pak četnost klesá a pohybuje se v rámci stovek, desítek a pak i jednotek zákazníků. Nejvyšší pak dosažená hodnota objednávek uskutečněná jediným zákazníkem je 43 objednávek za zmiňované období.

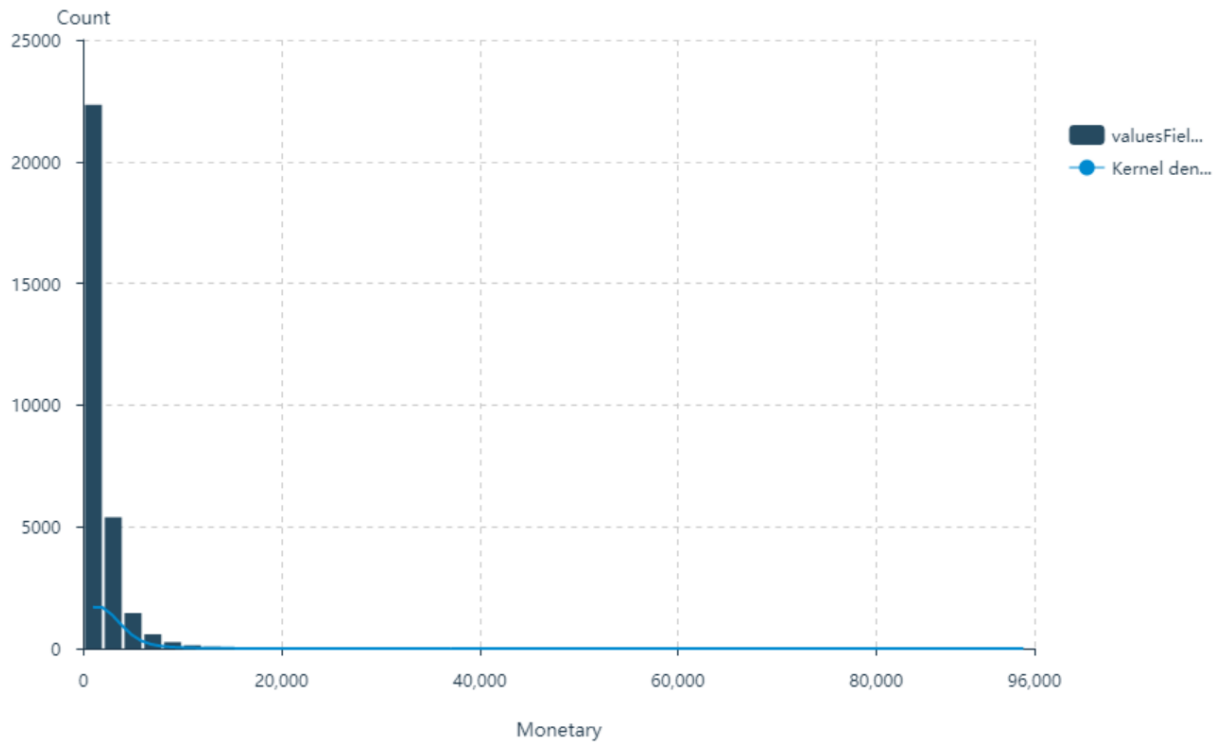


Obrázek 12 Histogram hodnoty F

Zdroj: vlastní zpracování

Histogram Monetary

Histogram hodnoty M (Peněžní hodnota) je znázorněna na následujícím obrázku 13. Na ose x jsou opět jednotlivé intervaly hodnoty, v tomto případě v intervalu po 2 000 korunách českých a na ose y je opět počet jednotlivých zákazníků dosahujících tuto hodnotu. Nejvíce zákazníků uskutečnilo objednávky za celkovou hodnotu prvního sloupce histogramu, a to v rozmezí 0 – 2 000 Kč, těchto zákazníků bylo celkem 22 340. Dalším nejčetnějším intervalem je hned interval následující, a to rozmezí 2 000 – 4000 Kč s četností 5 395 zákazníků. Tento histogram stejně jako u hodnoty F má sklon klesající, a proto s přibývajícím hodnotou objednávek počet zákazníků klesá. Nejvyšší utracenou částkou je 92 892,44 Kč v intervalu 92 000 – 94 000 Kč, kterou uskutečnil jediný zákazník.

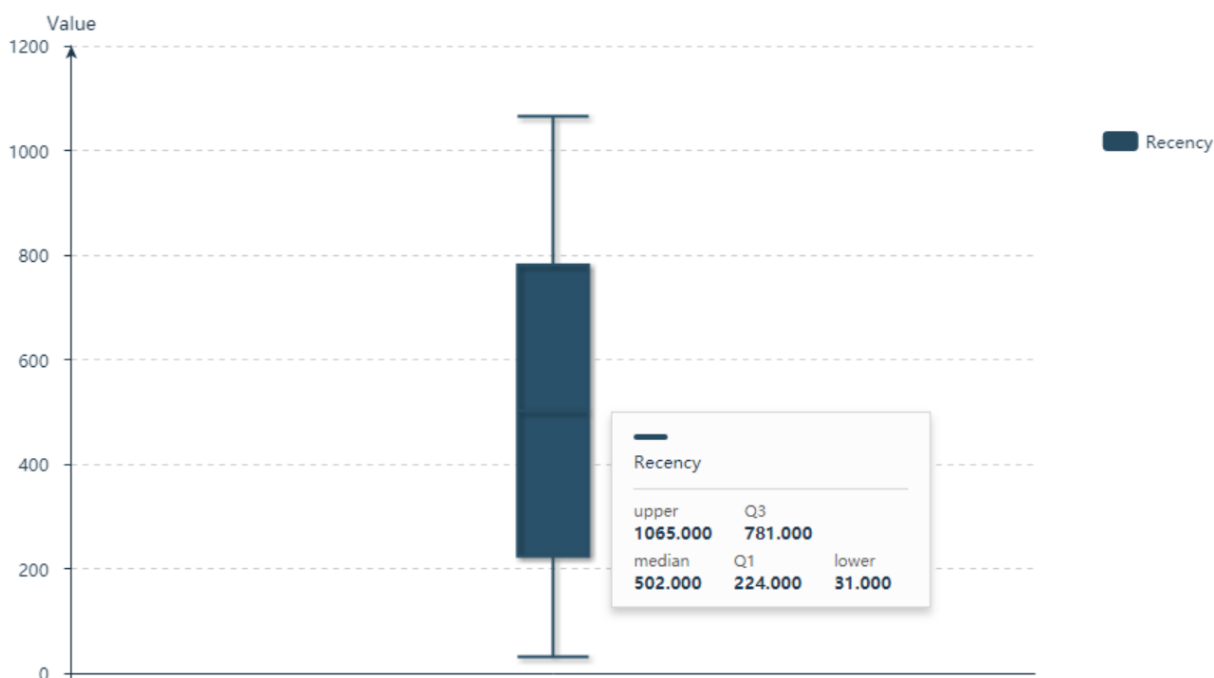


Obrázek 13 Histogram hodnoty M

Zdroj: vlastní zpracování

Krabicový graf Recency

Následující obrázek 14 znázorňuje krabicový graf hodnoty R (Aktuálnost), která podle grafu dosahuje minima v hodnotě 31 a maxima v hodnotě 1 065, tyto hodnoty jsou v jednotkách dní a vyznačují počet dní uplynulých od poslední objednávky daného zákazníka k datu provedení analýzy, v tomto případě tedy 1. 12. 2020. Hodnota značená jako Q1 je hranicí prvního kvartilu a nabývá hodnoty 224 dnů, dále následuje hodnota mediánu a to 502 dnů a jako poslední je zde hodnota Q3, která je hranicí třetího kvartilu a dosahuje hodnoty 781 dnů.

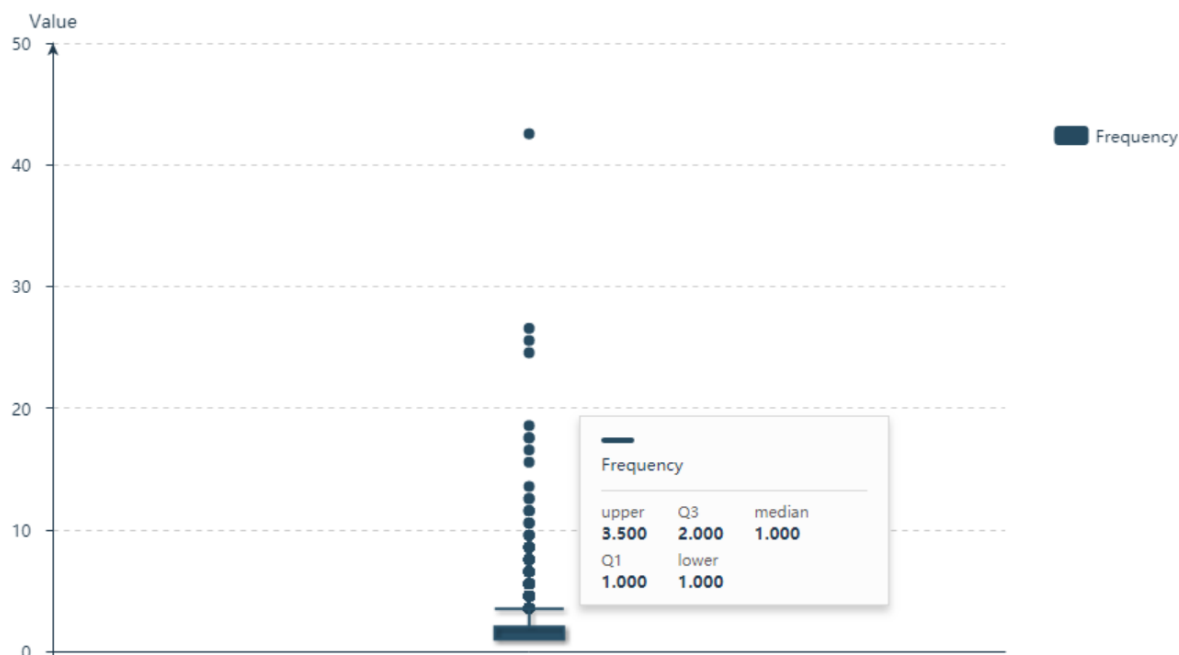


Obrázek 14 Krabicový graf hodnoty R

Zdroj: vlastní zpracování

Krabicový graf Frequency

Krabicový graf znázorněný na obrázku 15 zobrazuje hodnotu F (Četnost), ta dosahuje minima i zároveň hranici prvního kvartilu a mediánu v hodnotě 1 a maxima v hodnotě 3,5, v tomto případě jde o počty uskutečněných objednávek. Hranice pak třetího kvartilu je v hodnotě 2 objednávek. U této hodnoty dochází i k dosažení extrémních hodnot, které jsou vykresleny zvlášť jako samostatné body. Všechny tyto extrémní hodnoty se nacházejí na horním konci grafu a značí, že distribuce není symetrická, jak již bylo patrné u histogramu této hodnoty. Nejvyšší extrémní hodnotou je pak 43 objednávek.

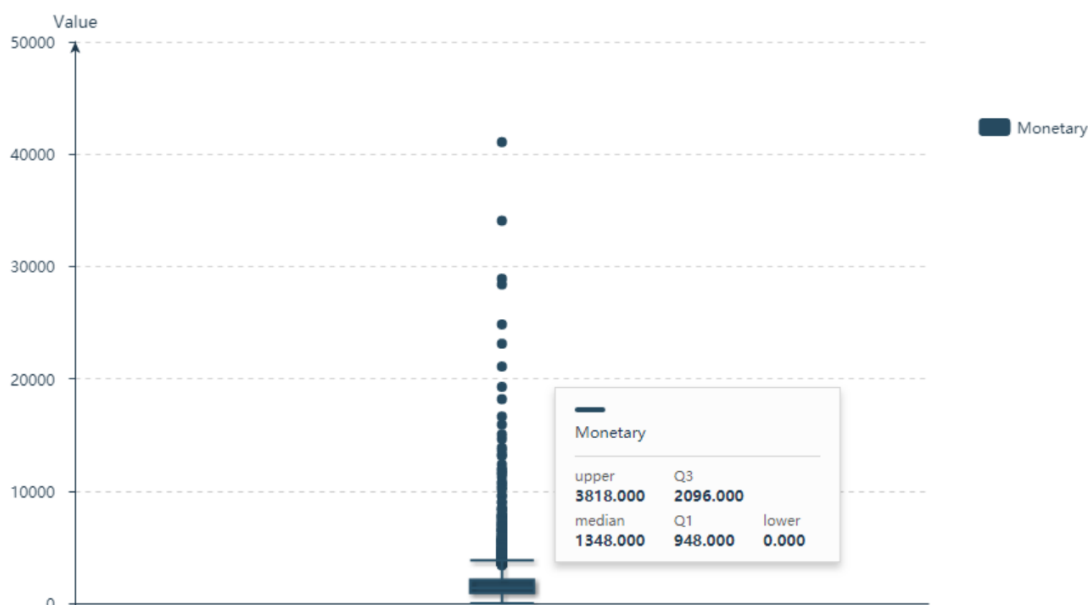


Obrázek 15 Krabicový graf hodnoty F

Zdroj: vlastní zpracování

Krabicový graf Monetary

Jako poslední znázorňovaný krabicový graf vyznačuje hodnotu M (Peněžní hodnota) viz. obrázek 16. Minimem této hodnoty je 0 a maximem 3 818, zde se jedná o jednotky korun českých. Hranice prvního kvartilu nabývá hodnoty 948 Kč, medián 1 348 Kč a hranice třetího kvartilu 2 096 Kč. Stejně jako u hodnoty F i zde se nacházejí extrémní hodnoty na horním konci grafu vyznačené samostatnými body a opět značí nesymetrickou distribuci hodnot. Nejvyšší extrém hodnoty M znázorněný v tomto grafu je 41 521,64 Kč, přičemž skutečně nejvyšší utracenou částkou jediným zákazníkem byla hodnota 92 892,44 Kč uvedena ve výše znázorněném histogramu této hodnoty.



Obrázek 16 Krabicový graf hodnoty M

Zdroj: vlastní zpracování

Dalším a velmi podstatným krokem je určení hraničních hodnot jednotlivých segmentů.

3.4.1 RFM analýza – přístup 1

Ve variantě určené softwarem budou jednotlivé hodnoty R, F a M rozděleny do 5 skupin s následujícími váhami: hodnota R – Aktuálnost má váhu 100, hodnota F – Četnost váhu 10 a M – Peněžní hodnota váhu 1. Tento počet skupin je ve výchozím nastavení softwaru, neboť se zákazníci nejčastěji pomocí RFM analýzy člení právě na 5 segmentů. Přestože bylo softwarem nastaveno 5 skupin pro každou hodnotu, software nakonec vytvořil pro hodnotu F skupiny jen 4.

Hraniční hodnoty byly softwarem určeny následovně:

Recency (Aktuálnost) – hodnoty jsou v jednotkách dnů, jelikož jsou poslední data dostupná k 31. 10. 2020, byl stanoven datum analýzy k 1. prosinci 2020, a proto je nejnovější objednávka uskutečněná před 31 dny.

Segment	Spodní hodnota	Horní hodnota
1	≥850	≤1 065
2	≥618	<850
3	≥364	<618

4	≥ 193	< 364
5	≥ 31	< 193

Frequency (Četnost) – hodnoty udávají počet uskutečněných objednávek. Software v tomto případě ne úplně ideálně určil jednotlivé intervaly, neboť podle histogramu hodnoty F uvedeného výše na obrázku 12 vidíme její silné zešikmení.

Segment	Spodní hodnota	Horní hodnota
1	≥ 1	≤ 1
2	> 1	≤ 2
3	> 2	≤ 3
4	> 3	≤ 43

Monetary (Peněžní hodnota) – tyto hodnoty pak tvoří hodnota objednávek.

Segment	Spodní hodnota	Horní hodnota
1	≥ 0	≤ 848
2	> 848	$\leq 1\ 148$
3	$> 1\ 148$	$\leq 1\ 548$
4	$> 1\ 548$	$\leq 2\ 444$
5	$> 2\ 444$	$\leq 95\ 708,82$

Toto rozdělení opět není ideální, což je patrné z vizualizace viz. Obrázek 13, který je silně zešikmen.

Po spuštění analýzy bylo ke každému zákazníkovi přiděleno jeho skóre R, F a M hodnoty a zároveň i celkové RFM skóre viz obrázek 17.

Table (8 fields, 30 443 records) #2

File Edit Generate

Table Annotations

	ID zákazníka	Recency	Frequency	Monetary	Recency Score	Frequency Score	Monetary Score	RFM Score
1	16091	636	1	748.000	2	1	1	211.000
2	25532	926	1	848.000	1	1	1	111.000
3	19253	722	1	1278.100	2	1	3	213.000
4	26323	940	1	1198.000	1	1	3	113.000
5	18865	708	1	748.000	2	1	1	211.000
6	4816	657	2	577.000	2	2	1	221.000
7	32635	1065	1	427.000	1	1	1	111.000
8	5442	200	2	3443.000	4	2	5	425.000
9	24063	117	1	798.000	5	1	1	511.000
10	5846	135	2	1365.000	5	2	3	523.000
11	17614	661	1	1498.000	2	1	3	213.000
12	1736	503	4	10739.0...	3	4	5	345.000
13	29538	196	1	698.000	4	1	1	411.000
14	30436	1024	1	586.000	1	1	1	111.000
15	2152	35	3	2679.070	5	3	5	535.000
16	9394	294	1	1597.000	4	1	4	414.000
17	24309	896	1	1099.000	1	1	2	112.000
18	9057	291	1	2047.000	4	1	4	414.000
19	14463	567	1	1847.000	3	1	4	314.000
20	14548	570	1	1248.000	3	1	3	313.000

Obrázek 17 RFM skóre – přístup 1

Zdroj: vlastní zpracování

Výsledná data byla vložena do 3D modelu pro jejich lepší vizualizaci. Na souřadnici x bylo přiděleno Recency (Aktuálnost) skóre, na souřadnici y Frequency (Četnost) skóre a na souřadnic z Monetary (Peněžní hodnota) skóre. Data pak byla vyjádřena podle celkového RFM skóre, které nabývalo hodnot od 111 do 545. Toto skóre bylo i rozděleno barevně, kdy modré zbarvení vyjadřuje nejnižší hodnoty RFM tedy 111, kdy zákazník nakoupil pouze jednou, před dlouhou dobou a za nízkou částku. Naopak tmavě růžové zbarvení ukazuje nejvyšší skóre a to 545, kdy daný zákazník nakoupil nedávno, už po několikáté a za vysokou částku. Velikost jednotlivých bodů v grafu je pak dána v závislosti na jednotlivá ID zákazníků, tedy jejich četnost. Výsledný 3D model je znázorněn na následujícím obrázku 18.



Obrázek 18 3D vizualizace – přístup 1

Zdroj: vlastní zpracování

Tento přístup, jak už bylo patrné u ne vhodně nastavených hraničních hodnot není nejideálnější a neposkytuje dobré výsledky, ale i tak může posloužit jako jedna z alternativ.

3.4.2 RFM analýza – přístup 2

U této varianty byl ponechán stejný počet skupin jako u varianty stanovené softwarem, tedy hodnoty R a M po 5 skupinách a hodnota F na 4 skupiny. Váhy byly rovněž ponechány stejné jako v předchozí variantě a to: hodnota R s váhou 100, hodnota F s váhou 10 a hodnota M s váhou 1. Hraniční hodnoty pak byly stanoveny následovně:

Recency (Aktuálnost) – jako výchozí zde byla považována hodnota mediánu, který činil 502 dní.

Segment	Spodní hodnota	Horní hodnota
1	≥750	≤1065
2	≥502	<750
3	≥210	<502
4	≥121	<210
5	≥31	<121

Frequency (Četnost) – u četnosti byly zanechány první segmenty podobné verzi softwaru, neboť nejhorší segment má právě charakteristický jen jeden nákup a na základě údajů byl průměrný počet na zákazníka jen 1,6 nákupu, proto tato hodnota nemůže být rozdělena na segmenty rovnoměrně. U vyšších segmentů už bylo rozpětí navyšováno, kdy poslední segment má rozpětí největší.

Segment	Spodní hodnota	Horní hodnota
1	≥ 1	≤ 1
2	> 1	≤ 3
3	> 3	≤ 15
4	> 15	≤ 43

Monetary (Peněžní hodnota) – zde bylo rozpětí určováno v závislosti na mediánu hodnot objednávek, který činil 1 348 Kč.

Segment	Spodní hodnota	Horní hodnota
1	≥ 0	≤ 995
2	> 995	$\leq 1\,348$
3	$> 1\,348$	$\leq 5\,995$
4	$> 5\,995$	$\leq 9\,995$
5	$> 9\,995$	$\leq 95\,708,82$

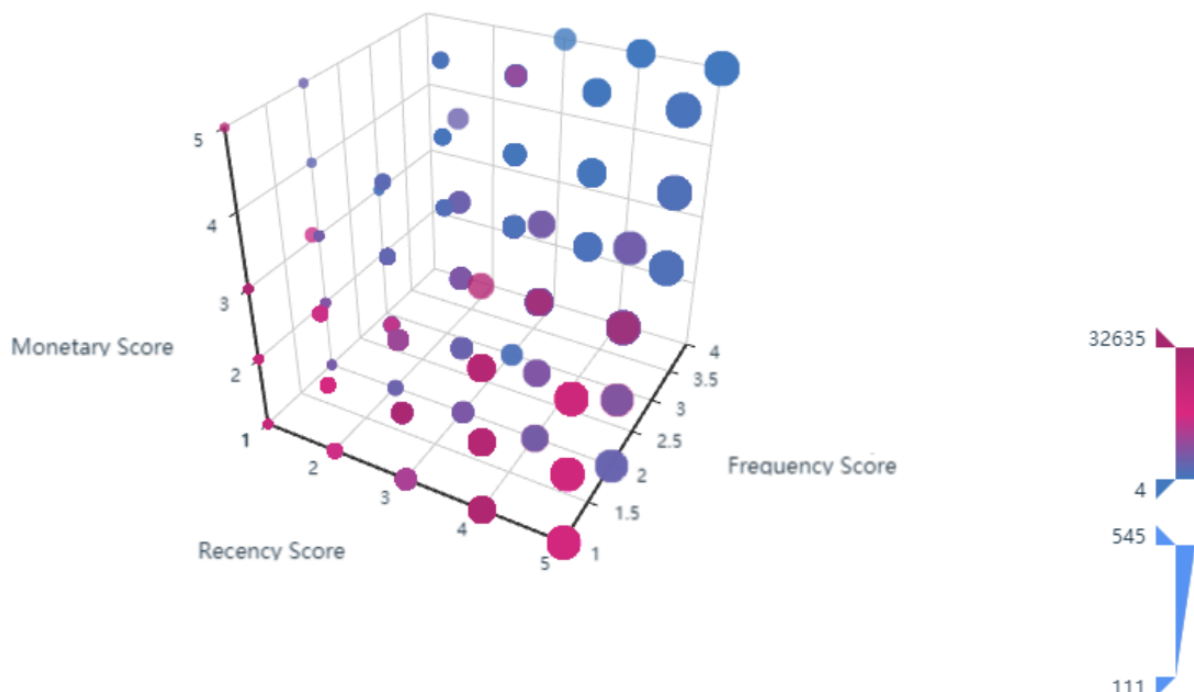
Po spuštění analýzy bylo ke každému zákazníkovi opět přiděleno jeho skóre R, F a M hodnoty a zároveň i celkové RFM skóre viz obrázek 19.

	ID zákazníka	Recency	Frequency	Monetary	Recency Score	Frequency Score	Monetary Score	RFM Score
1	16091	636	1	748.000	2	1	1	211.000
2	25532	926	1	848.000	1	1	1	111.000
3	19253	722	1	1278.100	2	1	2	212.000
4	26323	940	1	1198.000	1	1	2	112.000
5	18865	708	1	748.000	2	1	1	211.000
6	4816	657	2	577.000	2	2	1	221.000
7	32635	1065	1	427.000	1	1	1	111.000
8	5442	200	2	3443.000	4	2	3	423.000
9	24063	117	1	798.000	5	1	1	511.000
10	5846	135	2	1365.000	4	2	3	423.000
11	17614	661	1	1498.000	2	1	3	213.000
12	1736	503	4	10739.0...	2	3	5	235.000
13	29538	196	1	698.000	4	1	1	411.000
14	30436	1024	1	586.000	1	1	1	111.000
15	2152	35	3	2679.070	5	2	3	523.000
16	9394	294	1	1597.000	3	1	3	313.000
17	24309	896	1	1099.000	1	1	2	112.000
18	9057	291	1	2047.000	3	1	3	313.000
19	14463	567	1	1847.000	2	1	3	213.000
20	14548	570	1	1248.000	2	1	2	212.000

Obrázek 19 RFM skóre – přístup 2

Zdroj: vlastní zpracování

Výsledná data byla opět vložena do 3D modelu pro jejich lepší vizualizaci. Na souřadnice x, y a z byla přidělena stejná skóre jako v předchozí variantě. Data byla rovněž vyjádřena podle celkového RFM skóre, které nabývalo hodnot od 111 do 545. Výsledný 3D model je znázorněn na následujícím obrázku 20.



Obrázek 20 3D vizualizace – přístup 2

Zdroj: vlastní zpracování

Pro grafické a souvislé znázornění je na následujícím obrázku 21 zobrazen datový audit všech atributů.

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
ID zákazníka		Continuous	1	32636	16221.073	9453.228	0.014	--	30443
Recency		Continuous	31	1065	510.838	312.822	0.167	--	30443
Frequency		Continuous	1	97	1.624	1.808	12.630	--	30443
Monetary		Continuous	0.000	95708.820	1989.905	2615.158	10.192	--	30443
Recency Score		Nominal	1	5	--	--	--	5	30443
Frequency Sc...		Nominal	1	4	--	--	--	4	30443
Monetary Score		Nominal	1	5	--	--	--	5	30443
RFM Score		Continuous	111.000	545.000	317.472	144.553	0.011	--	30443

Obrázek 21 Datový audit

Zdroj: vlastní zpracování

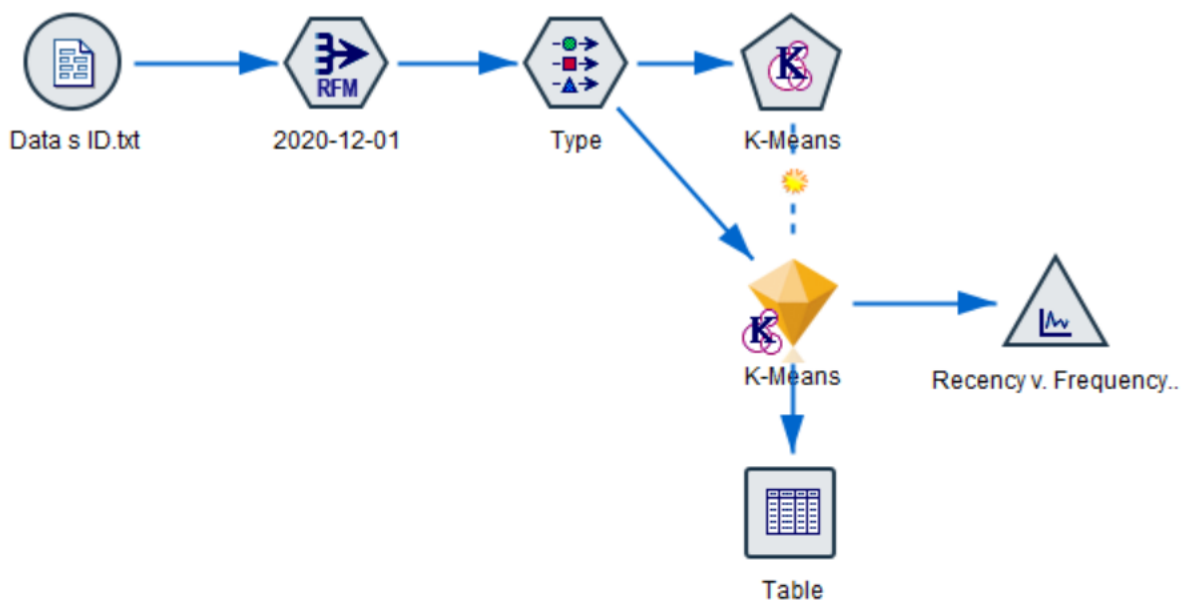
3.4.3 RFM analýza – přístup 3

V následujícím přístupu bude využit uzel K-Means rovněž dostupný v prostředí IBM SPSS Modeler.

Uzel K-Means

Uzel K-Means poskytuje metodu klastrové analýzy. Záznamy jsou pomocí tohoto uzlu seskupeny tak, aby si byly v jednotlivých klastrech navzájem podobné, ale k záznamům v dalších klastrech odlišné (IBM Documentation, 2021c).

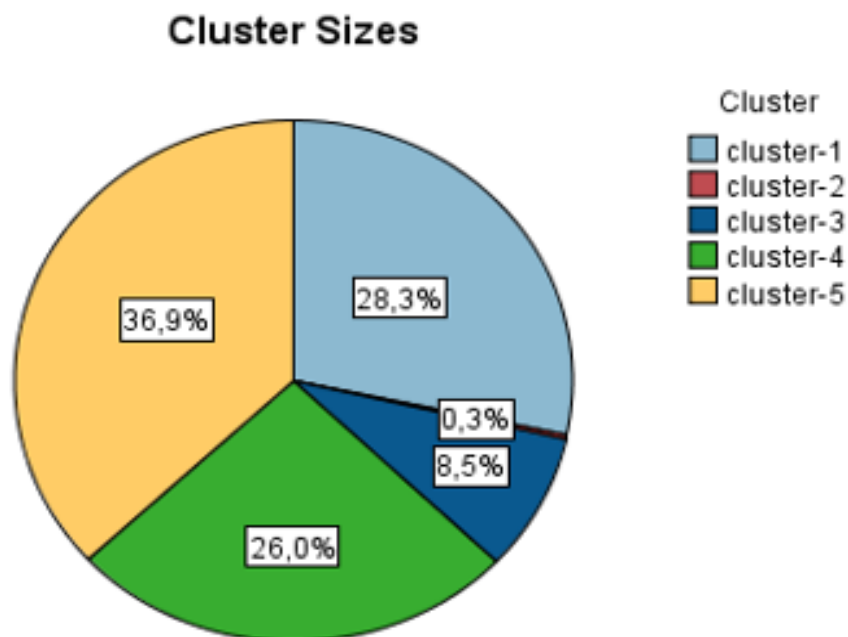
Na obrázku 22 je znázorněn stream, pomocí něhož bylo dosaženo rozdělení dat do jednotlivých klastrů. Jako vstupní atributy byly použity hodnoty R, F a M stejně jako u prvních dvou přístupů v uzlu RFM Analysis.



Obrázek 22 K-Means stream

Zdroj: vlastní zpracování

Po spuštění výše znázorněného streamu byla data rozdělena do 5 klastrů, které byly znázorněny pomocí výsečového grafu na obrázku 23. Největší klastr neboli klastr 5 odpovídá 36,9 %, což je 11 227 zákazníků, oproti tomu nejmenší klastr 2 je tvořen pouhými 0,3 % zákazníků (tj. 99 zákazníků).



Obrázek 23 Procentní znázornění klastrů

Zdroj: vlastní zpracování

Jednotlivým klastrům byla vždy určena průměrná hodnota všech vstupních atributů R, F a M. Na následujícím obrázku 24 je vždy uvedeno označení klastru, procentní velikost s počtem zákazníků v něm a jednotlivé průměrné hodnoty atributů. Klastry jsou zde seřazeny podle velikosti neboli četnosti zákazníků v nich. Již zmiňovaný největší klastr 5 má pak průměrnou hodnotu R 230,96 dne, průměrnou hodnotu F 1,35 objednávky a průměrnou hodnotu M 1 529,87 Kč. Oproti tomu nejmenší klastr 2 má průměrnou hodnotu R 132,52 dne, průměrnou hodnotou F 19,16 objednávky a průměrnou hodnotu M 30 170,38 Kč.

Cluster	cluster-5	cluster-1	cluster-4	cluster-3	cluster-2
Label					
Description					
Size	36,9% (11227)	28,3% (8622)	26,0% (7905)	8,5% (2590)	0,3% (99)
Inputs	Frequency 1,35	Frequency 1,40	Frequency 1,14	Frequency 4,38	Frequency 19,16
	Monetary 1 529,87	Monetary 1 700,79	Monetary 1 382,87	Monetary 5 722,10	Monetary 30 170,38
	Recency 230,96	Recency 606,55	Recency 928,55	Recency 144,95	Recency 132,52

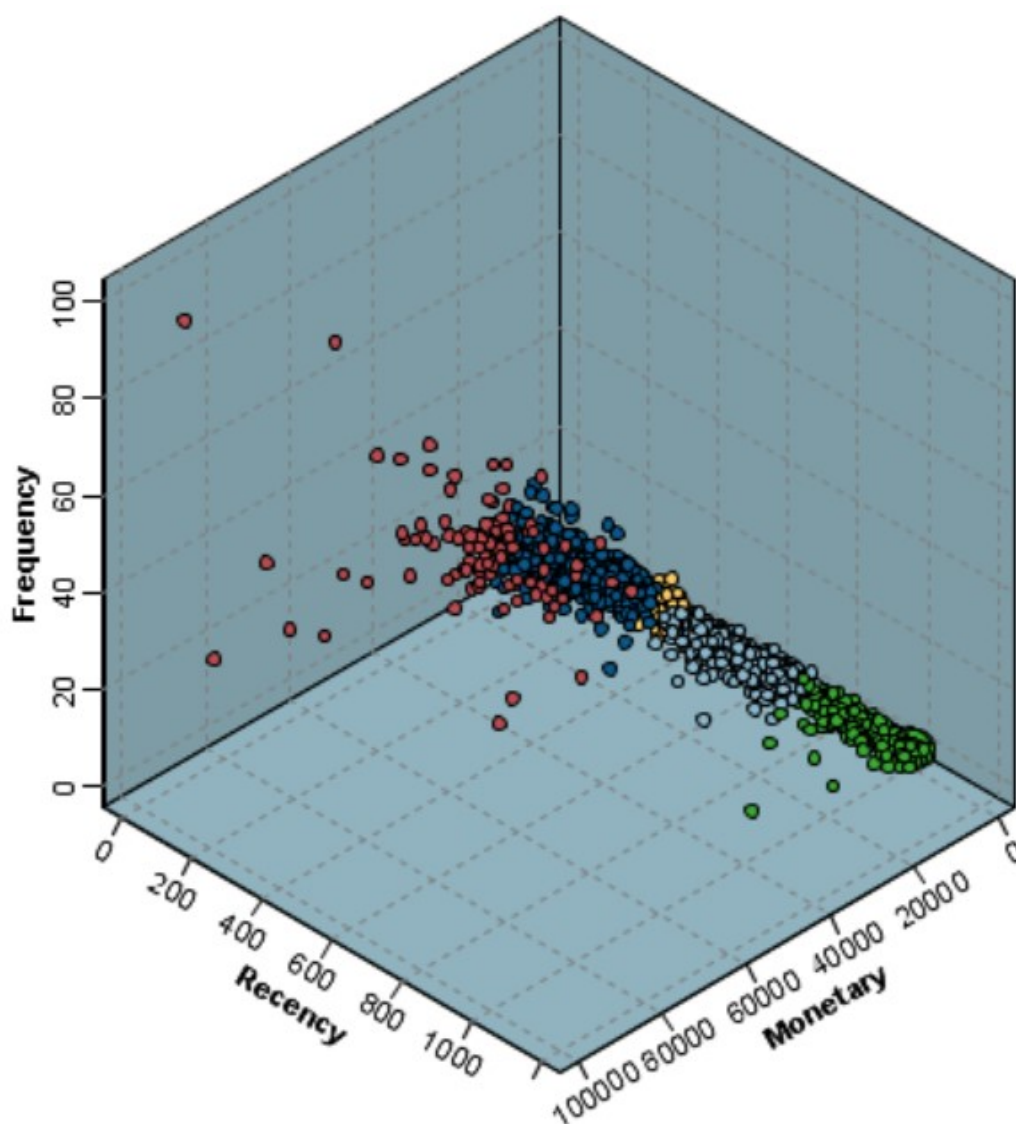
Obrázek 24 Průměrné hodnoty klastrů

Zdroj: vlastní zpracování

Pro společnost je nejzajímavější klastr 2, neboť dosahuje nejnižší průměrné hodnoty R neboli nejméně dnů od poslední objednávky (tj. 132,52), nejvyšší průměrné hodnoty F neboli počtu objednávek na zákazníka (tj. 19,16) a nejvyšší průměrné hodnoty M neboli celkové hodnoty objednávek (tj. 30 170,38). Naopak nejméně zajímavým klastrem je klastr 4. Klastr 4

dosahuje nejvyšší hodnoty R, a to v průměru 928,55 dnů od poslední objednávky, nejnižší hodnoty F, která činí průměrně 1,14 objednávky na zákazníka a nejnižší hodnoty M, která nabývá v průměru 1 382,87 Kč. Druhým nejlepším klastrem je klastr 3, který má hned po klastru 2 nejlepší hodnoty R, F ale i M. Co se týče klastrů 1 a 5, tak ty jsou si podobné jak procentním zastoupením, tak i hodnotami F a M, v hodnotě R se ale už rozcházejí, kdy klastr 1 má téměř 3x větší průměrnou hodnotu R než klastr 5.

Na základě těchto dat byla možnost vytvořit opět 3D vizualizaci, která je znázorněna na obrázku 25. Jednotlivá data klastrů jsou odlišena barevně, data klastru 1 jsou barvy světle modré, klastru 2 barvy červené, klastru 3 barvy tmavě modré, klastru 4 barvy zelené a klastru 5 barvy žluté.



Obrázek 25 3D vizualizace – přístup 3

Zdroj: vlastní zpracování

V rámci shlukování je používán koeficient siluety, který vyjadřuje, jak je daný model shlukování dobrý, neboť jak dobře jsou jednotlivé shluky sestaveny. V tomto případě vyšel koeficient siluety v hodnotě 0,5 (znázorněno na obrázku 26), což značí v celku dobrý výsledek a model je tak vhodně sestaven.



Obrázek 26 Koeficient siluety

Zdroj: vlastní zpracování

V rámci streamu byl taktéž využit uzel Table, který vyjádřil jednotlivé údaje pro každého zákazníka na základě jeho ID. V tomto výstupu, jehož ukázka 20 zákazníků je uvedena na obrázku 27, jsou pro každé ID vyjádřeny jednotlivé hodnoty R, F a M, ale i do jakého klastru daný zákazník spadá.

	ID zákazníka	Recency	Frequency	Monetary	\$KM-K-Means
1	16091	636	1	748.000	cluster-1
2	25532	926	1	848.000	cluster-4
3	19253	722	1	1278.100	cluster-1
4	26323	940	1	1198.000	cluster-4
5	18865	708	1	748.000	cluster-1
6	4816	657	2	577.000	cluster-1
7	32635	1065	1	427.000	cluster-4
8	5442	200	2	3443.000	cluster-3
9	24063	117	1	798.000	cluster-5
10	5846	135	2	1365.000	cluster-5
11	17614	661	1	1498.000	cluster-1
12	1736	503	4	10739.0...	cluster-1
13	29538	196	1	698.000	cluster-5
14	30436	1024	1	586.000	cluster-4
15	2152	35	3	2679.070	cluster-3
16	9394	294	1	1597.000	cluster-5
17	24309	896	1	1099.000	cluster-4
18	9057	291	1	2047.000	cluster-5
19	14463	567	1	1847.000	cluster-1
20	14548	570	1	1248.000	cluster-1

Obrázek 27 Zařazení jednotlivých ID do klastrů

Zdroj: vlastní zpracování

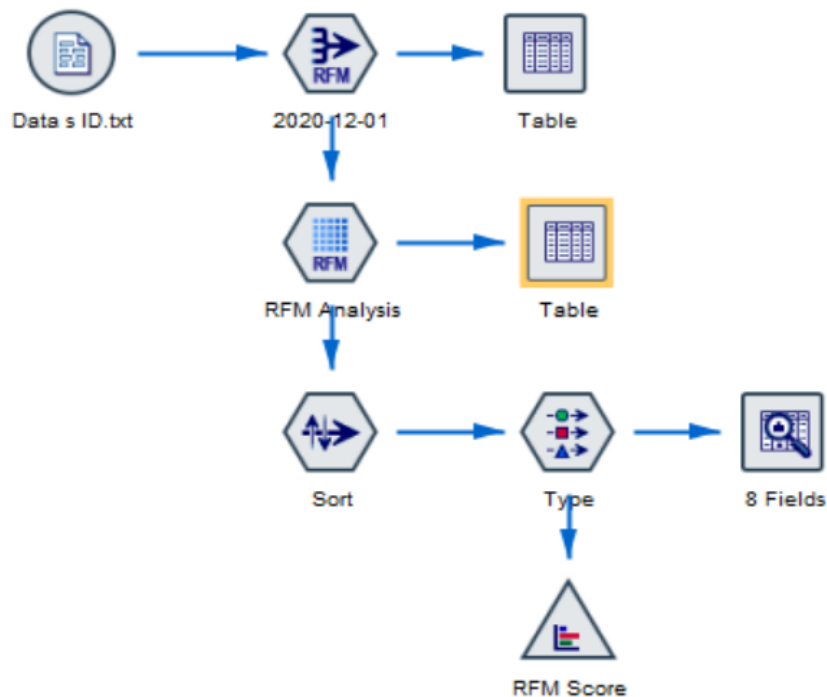
4 VYHODNOCENÍ VÝSLEDKŮ A FORMULACE ZÁVĚRŮ

V následující kapitole bude celkové shrnutí přístupu 2 a 3 provedené RFM analýzy na dostupných datech a rozdělení zákazníků na segmenty. Co se týče přístupu 1, ten nebyl softwarem ideálně nastaven kvůli hraničním hodnotám. Jelikož metody přístupů 2 a 3 jsou rozdílné, byly použity i rozdílné přístupy segmentace zákazníků, v tomto případě pak ale nelze porovnat zařazení do segmentů podle použitých metod.

4.1 Vyhodnocení přístupu 2

U tohoto přístupu byly využity stejně jako u přístupu 1 uzly RFM Aggregate a RFM Analysis v prostředí IBM SPSS Modeler. U hraničních hodnot byla využita možnost nastavení dle vlastně určených hodnot, které byly vybrány na základě mediánů R, F a M hodnot.

V předchozí kapitole byla představena samotná analýza na dostupných datech a následná 3D vizualizace. Pro lepší představu byl ale za účelem vyhodnocení dodán do streamu uzel distribučního grafu, který znázorňuje jednotlivé četnosti všech hodnot RFM skóre, tedy od 111 do 545. Doplněný stream a následná ukázka výstupu distribučního grafu jsou znázorněny na obrázcích 28 a 29.



Obrázek 28 Doplněný stream o distribuční graf – přístup 2

Zdroj: vlastní zpracování

Value /	Proportion	%	Count
111.000		9.16	2790
112.000		7.27	2213
113.000		7.6	2313
114.000		0.03	9
115.000		0.01	2
121.000		0.07	22
122.000		0.19	57
123.000		2.53	769
124.000		0.05	14
125.000		0.01	3
133.000		0.08	24
134.000		0.03	10
135.000		0.01	4
211.000		6.72	2047
212.000		5.4	1645
213.000		5.28	1608
214.000		0.02	7
215.000		0.01	3

Obrázek 29 Ukázka části distribučního grafu s RFM skóre

Zdroj: vlastní zpracování

Na základě distribučního grafu bylo zjištěno, že největší četnosti zákazníků nabývá skóre 111 s 9,16 % zákazníků (tj. 2790). Toto skóre je bohužel pro společnost nejméně zajímavé, neboť tito zákazníci uskutečnili jen jednu objednávku před dlouhou dobou a za nízkou částku. V tomto případě se jedná o objednávku uskutečněnou před více než 25 měsíci od data provedení analýzy, tzn. objednávka uskutečněná v roce 2018 a v cenovém rozmezí od 0 – 995 Kč. Naopak nejlepší možné dosažené skóre 545 dosáhlo pouhých 0,15 % zákazníků, což je 46 zákazníků. Tito zákazníci uskutečnili nejnovější objednávku nejdéle 4 měsíce před datem provedení analýzy, tzn. od srpna 2020 do konce října 2020. Jejich F hodnota značila počet objednávek v rozmezí 15 – 43 a za všechny tyto objednávky utratili částky v intervalu 9 995 – 95 708,82 Kč. Přičemž zmiňovaná hraniční hodnota 95 708,82 Kč byla uskutečněna pouze jedním zákazníkem.

Nyní budou zákazníci rozděleni do segmentů uvedených v kapitole 1 v přístupu podle Štráfěldy. Pro připomenutí se jedná o segmenty následující:

- Noví zákazníci,
- Zlatí zákazníci,
- Loajální zákazníci,
- Upadající zákazníci,

- Ztracení zákazníci.

Do segmentu **Noví zákazníci** spadají zákazníci, kteří nakoupili v nedávné době a pouze jednou. Sem tedy byli zařazeni zákazníci, kteří u hodnoty R dosahovali skóre 4 a 5, u hodnoty F skóre 1 a u hodnoty M skóre 1-5. Po sečtení četností všech těchto skóre bylo určeno tomuto segmentu bez mála 4 000 zákazníků (tj. cca 13 %). Této skupině zákazníků je doporučeno zasílání nabídky formou newsletteru či mailingu, který může být nastaven k pravidelnému zasílání po uplynutí určité doby od poslední objednávky.

Zlatí zákazníci jsou pak ti zákazníci, kteří nakupují často, za vysoké částky a naposledy v nedávné době. Skóre pak jednotlivých hodnot bylo následující: hodnota R skóre 4 a 5, hodnota F skóre 3 a 4 a hodnota M skóre 4 a 5. Tento segment je ze všech nejmenší, neboť zákazníků nabývajících všechny tyto hodnoty zároveň je pouze 647, což jsou pouhé 2 % zákazníků. Tito zákazníci jsou pro společnost nejvíce důležití. Společnost by mohla na tento segment aplikovat osobitější přístup, vytvoření například nějakého VIP klubu či je využít pro šíření pozitivních recenzí.

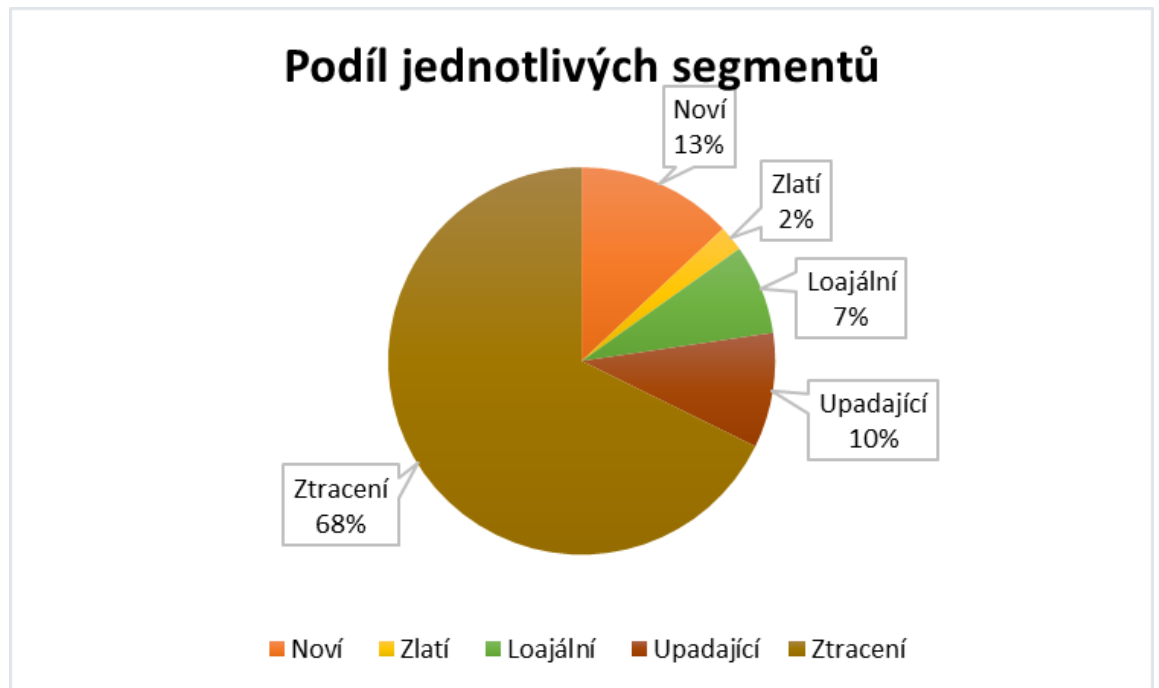
Mezi **Loajální zákazníky** spadají ti, kteří nakupují pravidelně a v nedávné době jako zákazníci zlatí, ale už ne za tak vysoké částky. Tomuto segmentu pak byla určena následující skóre: u hodnoty R skóre 4 a 5, u hodnoty F skóre 2 a 3 a hodnota M skóre 1-3. Tento segment pak dosahoval cca 7 % zákazníků, (tj. 2 287). U těchto zákazníků je doporučena možnost věrnostního programu, kdy může společnost ještě více posílit jejich loajalitu.

Do segmentu **Upadajících zákazníků** patří zákazníci, kteří dříve u společnosti nakupovali, ale v poslední době objednávku neuskutečnili. Do tohoto segmentu byli začleněni zákazníci s následujícími skóre hodnot: u hodnoty R skóre 3, u hodnoty F skóre 2-4 a u hodnoty M skóre 1-5. Ve výsledku patří do tohoto segmentu téměř 10 % zákazníků (tj. 2 920). Společnost by se měla pokusit zákazníka přimět uskutečnit novou objednávku, neboť jinak může o zákazníka definitivně přijít. Variantou, jak nové objednávky dosáhnout pak může být nějaká výrazná sleva či časově omezená nabídka.

Posledním segmentem a bohužel nejvíce zastoupeným je segment **Ztracených zákazníků**. V tomto případě tvoří tento segment naprostá většina zákazníků, a to bez mála 68 % (tj. 20 618). Tito zákazníci už po velmi dlouhou dobu žádnou objednávku neuskutečnili, a proto jim bylo přiděleno skóre hodnoty R 1-2. Co se týče hodnoty F, tam bylo stanoveno skóre 1-3 a u hodnoty M skóre 1-5. Do tohoto segmentu spadají i zákazníci, kteří byli dříve zákazníky zlatými či loajálními. V tomto případě by se společnost mohla pokusit o jejich oživení stejně

jako u segmentu upadajících zákazníků. Pokud ale tomu tak není, je lepší se těchto zákazníků vzdát a investovat čas a prostředky do segmentů jiných.

Na následujícím obrázku 30 je pak pomocí výsečového grafu znázorněn podíl jednotlivých segmentů.

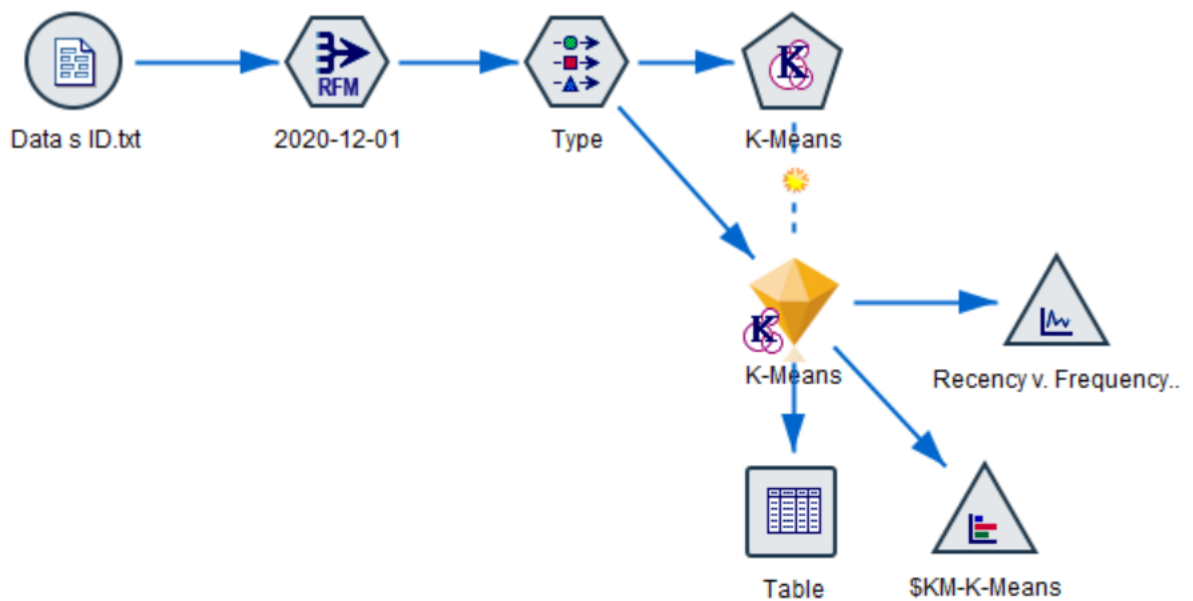


Obrázek 30 Podíl jednotlivých segmentů

Zdroj: vlastní zpracování

4.2 Vyhodnocení přístupu 3

U přístupu 3 byl využit uzel K-Means opět v prostředí IBM SPSS Modeler. Jak již bylo zmíněno v předchozích kapitolách, metoda K-Means se zabývá shlukováním dat do klastrů, kdy data v daných klastrech jsou si podobná, ale od dat z dalších klastrů odlišná. Do původního streamu znázorněném na obrázku 22 byl opět přidán uzel distribučního grafu pro lepší a srozumitelnější vizualizaci viz. obrázek 31.



Obrázek 31 Doplněný stream o distribuční graf – přístup 3

Zdroj: vlastní zpracování

Následný distribuční graf popisující četnosti a podíly jednotlivých klastrů je uveden na následujícím obrázku 32.

Value /	Proportion	%	Count
cluster-1		28.32	8622
cluster-2		0.33	99
cluster-3		8.51	2590
cluster-4		25.97	7905
cluster-5		36.88	11227

Obrázek 32 Distribuční graf klastrů

Zdroj: vlastní zpracování

Nyní budou zákazníci rozděleni do jednotlivých fází zákazníků, které uvádí Gary Seitz, zmíněné již v kapitole 1. Pro připomenutí se jedná o fáze následující:

- Fáze 1 – Zkoušeči,
- Fáze 2 – Kupující,
- Fáze 3 – Vysoká hodnota, klesající potenciál,
- Fáze 4 – Nízká hodnota, klesající potenciál.

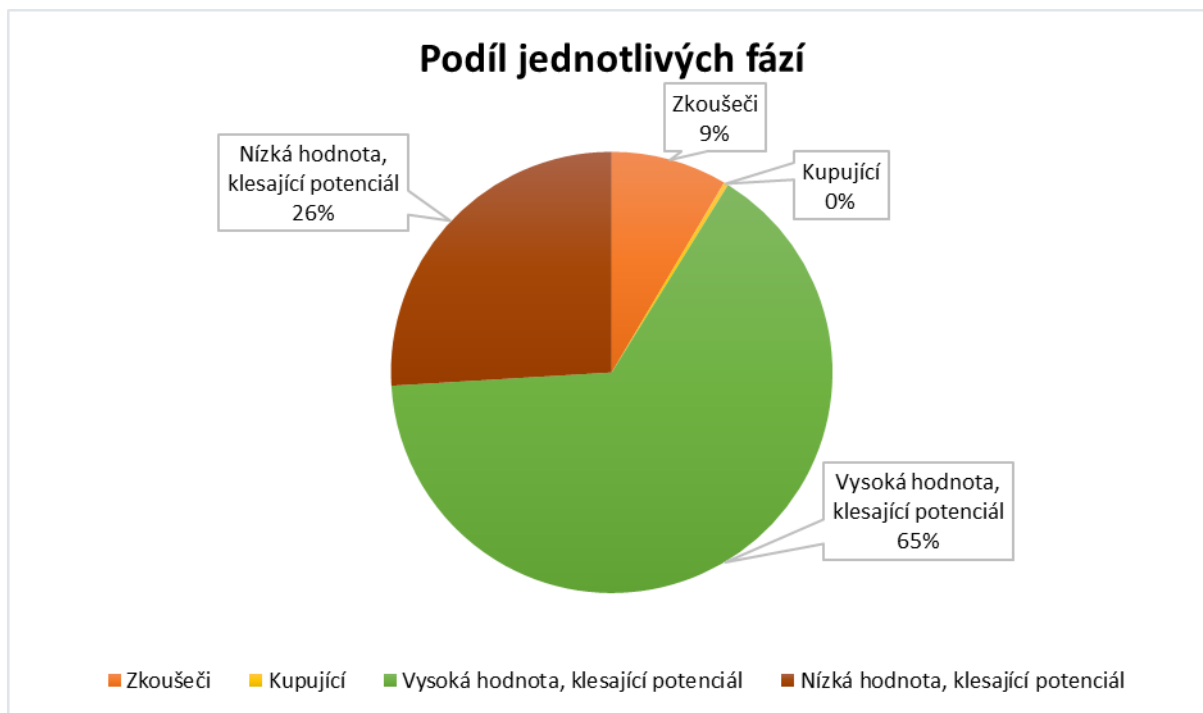
Do fáze **Zkoušeči** byl zařazen klastř 3, neboť se jedná o poměrně nové zákazníky, kteří nakoupili nedávno. Všechny 3 hodnoty R, F a M mají druhou nejlepší průměrnou hodnotu ze všech klastřů, a to R 144,95 dnů, F 4,38 objednávky a M 5 772,10 Kč. Tento klastř pak tvoří 8,51 % (tj. 2 590) zákazníků. Společnost by se měla snažit tyto zákazníky přimět, aby nadále nakupovali. Zákazníci v této fázi dobře reagují na marketing týkající se nejprodávanějších a nejnovějších produktů. Hlavním cílem je ale pro společnost dostat je do následující fáze 2.

Fází **Kupující** rozumíme ty nejlepší zákazníky společnosti, kterých chce mít společnost nejvíce, neboť dobře reagují na marketing a jsou nejlepším aktivem společnosti. Do této fáze byl zařazen klastř 2, který je ze všech nejlepší, neboť dosahuje nejlepších průměrných hodnot u všech 3 proměnných. Průměrná hodnota R je u tohoto klastř 132,52 dnů, hodnota F činí 19,16 objednávky a hodnota M v průměru 30 170,38 Kč. Bohužel v tomto případě je tento klastř nejmenší a tvoří ho pouze 0,33 %, což je pouhých 99 zákazníků. Cílem je pro společnost tyto zákazníky v dané fázi udržet.

Jelikož bylo pomocí uzlu K-Means vygenerováno 5 klastřů a fáze jsou pouze 4, rozhodla jsem se do následující fáze **Vysoká hodnota, klesající potenciál** zařadit 2 klastř najednou, a to klastř 1 a 5. Tyto klastř jsou si velmi podobné, jediné, v čem se rozcházejí je rozdílnější hodnota R. Tato fáze je ale charakteristická nenakoupením po delší dobu, ale stále slušnými hodnotami nákupu. Jelikož oba klastř nabývají druhou a třetí nejhorší hodnotou R, mohou být tak spojeni spolu. Ve výsledku se tak jedná o nejčtenější fázi zákazníků, neboť oba klastř tvoří dohromady 65,2 % (tj. 19 849) zákazníků. Společnost by se měla snažit o znovu zaktivování těchto zákazníků a přimět je k nákupu.

Poslední fázi a zároveň nejhorší pro společnost je fáze **Nízká hodnota, klesající potenciál**. Do této fáze byl zařazen klastř 4, který nabývá nejhorších hodnot u všech 3 proměnných. Průměrná hodnota R nabývá 928,55 dne od poslední objednávky, průměrný počet objednávek na zákazníka je pouhých 1,14 a průměrná hodnota objednávek je 1 382,87 Kč. Tento klastř tvoří 25,97 % (tj. 7 905) zákazníků. Pro společnost představuje tento klastř ztracené zákazníky, proto by měla přestat zbytečně vynakládat výdaje na tento segment a zaměřit se na lukrativnější a slibnější zákazníky.

Na následujícím obrázku 33 je pak pomocí výsečového grafu znázorněn podíl jednotlivých fází zákazníků.



Obrázek 33 Podíl jednotlivých fází zákazníků

Zdroj: vlastní zpracování

ZÁVĚR

Cílem mé diplomové práce bylo navrhnout použití RFM analýzy na dostupných datech.

RFM analýza je poměrně často používanou marketingovou analýzou pro segmentaci zákazníků. Hlavním důvodem jejího častého použití je její poměrná jednoduchost. Společnosti totiž k jejímu provedení nepotřebují značné znalosti či dovednosti, ale především základní údaje o uskutečněných objednávkách a popřípadě nějaký specializovaný software. Základní výpočty lze však provést například i v prostředí MS Excel.

Pro svou práci jsem se rozhodla využít prostředí softwaru IBM SPSS Modeler. Práce s tímto softwarem byla pro mě složitější, neboť jsem se do té doby s tímto softwarem nesetkala. IBM SPSS Modeler se ale prezentuje jednoduchým intuitivním ovládáním bez nutnosti předchozích znalostí či zkušeností a ve výsledku mě práce s ním bavila a obohatila o další zkušenosti.

V první části práce byly vysvětleny základní pojmy týkající se dané oblasti a jejich využití. Poté následovalo uvedení hlavního cíle diplomové práce spolu s představením společnosti poskytující data a jejich prvotní popis a vizualizace. Následovala úprava a zpracování dat ve zmiňovaném softwaru zakončené vyhodnocením výsledků a formulací závěrů.

Pro provedení analýzy na poskytnutých datech byly uskutečněny různé přístupy zpracování RFM analýzy. Prvním z nich byl pomocí uzlů RFM Analysis a RFM Aggregate, další byl pak pomocí uzlu K-Means. Z obou přístupů je patrné, že velká většina zákazníků dané společnosti uskutečnila pouze jeden nákup a většinou i před dlouhou dobou, tyto zákazníci jsou bohužel pro společnost ztracenými a nevyplatí se do nich dále investovat a směřovat na ně dál přímý marketing. Na druhou stranu jsou zde i zákazníci, kteří sice zatím nepatří mezi ty nejlepší, ale mají k tomu potenciál či tvoří stabilní a loajální zákaznickou základnu. Samozřejmě se zde i našli zákazníci spadající do nejlépe hodnocených skupin nakupujících velmi často, v nedávné době a za vysoké částky, těch je ale bohužel jen pár procent ze všech dosavadních zákazníků po dobu, ze které byla poskytnuta data.

Společnost by se měla více zaměřit na nějaký věrnostní klub či různé benefity pramenící se z opakovaných nákupů pro zvýšení procentního zastoupení lukrativnější části zákazníků, než je tomu nyní. Společnost v současné době dělí zákazníky jen na 2 segmenty, a to na „nové“ a „vracející se“. V poslední kapitole byly tedy přestaveny dvě různá rozdělení na více segmentů s návrhy, jak s daným segmentem komunikovat.

Pro svou diplomovou práci jsem čerpala převážně ze zahraničních odborných článků, neboť v české literatuře je bohužel RFM analýza zmiňována velmi okrajově.

Na závěr bych chtěla dané společnosti velmi poděkovat za možnost využití jejich dat pro zpracování této diplomové práce. Zpracování diplomové práce právě na toto téma bylo pro mě velkým přínosem a domnívám se, že její hlavní cíl byl naplněn.

POUŽITÁ LITERATURA

1. AGGELIS, Vasilis; CHRISTODOULAKIS, Dimitris, 2005. Customer clustering using rfm analysis. In: *Proceedings of the 9th WSEAS International Conference on Computers*.
2. BHOLOWALIA, Purnima; KUMAR, Arvind, 2014. EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*.
3. BIRANT, Derya, 2011. Data mining using RFM analysis. In: *Knowledge-oriented applications in data mining*. IntechOpen.
4. GUSTRIANSYAH, Rendra; SUHANDI, Nazori; ANTONY, Fery, 2020. Clustering optimization in RFM analysis based on k-means. *Indones. J. Electr. Eng. Comput. Sci.*
5. IBM [online], 2021. United States, [cit. 2021-5-13]. Dostupné z: <https://www.ibm.com/products/spss-modeler>
6. IBM Documentation [online], 2021a. United States, [cit. 2021-6-7]. Dostupné z: <https://www.ibm.com/docs/cs/spss-modeler/18.2.1?topic=nodes-rfm-aggregate-node>
7. IBM Documentation [online], 2021b. United States, [cit. 2021-6-7]. Dostupné z: <https://www.ibm.com/docs/cs/spss-modeler/18.2.1?topic=nodes-rfm-analysis-node>
8. IBM Documentation [online], 2021c. United States, [cit. 2021-6-14]. Dostupné z: <https://www.ibm.com/docs/cs/spss-modeler/18.2.1?topic=models-means-node>
9. Interní zdroje – e-mailová komunikace s ředitelkou společnosti, 2020
10. KODINARIYA, Trupti M.; MAKWANA, Prashant R., 2013. Review on determining number of Cluster in K-Means Clustering. *International Journal*.
11. KOTLER, Philip a Kevin Lane KELLER, 2007. *Marketing management*. Praha: Grada. ISBN 8024713594.
12. LOŠŤÁKOVÁ, Hana, et al., 2017. *Nástroje posilování vztahů se zákazníky na B2B trhu*. Grada Publishing as. ISBN 8027197988.
13. MIGLAUTSCH, J., 2000. Thoughts on RFM scoring. *Journal of Database Marketing*, č. 8, s. 67–72

14. NOVO, Jim, 2000. Customer Response, Retention and Valuation Concepts (RFM Model), [online] [vid. 13. únor 2021]. Dostupné z: <http://www.jimnovo.com/RFMtour.htm>
15. *Online marketing*. Brno: Computer Press, 2014. ISBN 9788025141557
16. RUD, Olivia Parr, Ivo MAGERA a Milan DANĚK, 2001. *Data mining: praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníků (CRM)*. Praha: Computer Press. Databáze. Rychle a jistě. ISBN 80-7226-577-6.
17. SEITZ, Gary, 2014. RFM: A Cool Tool for Simple Analytics. In: [online]. B.m. [vid. 13. únor 2021]. Dostupné z: <http://www.slideshare.net/CTRAC/rfm-2014>
18. ŠTRÁFELDA, Jan, 2021. RFM segmentace. *Štráfelda* [online] [cit. 2021-5-20]. Dostupné z: <https://www.strafelda.cz/rfm-segmentace>
19. WEI, Jo-Ting, et al., 2011. Using RFM model in a Children's dental clinic for market segmentation. In: *The 11th International Conference of Decision Science Institute and the 16th Annual Conference of Asia-Pacific Decision Science Institute, IDSI, Taipei, Taiwan*.
20. WEI, Jo-Ting; LIN, Shih-Yen; WU, Hsin-Hung, 2010. A review of the application of RFM model. *African Journal of Business Management*, 4.19: 4199-4206.
21. WIRTZ, Jochen, Patricia CHEW a Christopher H. LOVELOCK, 2017. *Essentials of Services Marketing*. 3 ed. Velká Británie: Pearson Education Limited. ISBN 9781292089959.