

Segmentace zákazníků nákupního centra na základě příjmu a nákupního chování

1. Úvod

Maloobchodní prodej a chování spotřebitelů procházejí v posledních desetiletích výraznou transformací (Kotler & Keller, 2016). Provozovatelé nákupních center již nemohou spoléhat na plošné marketingové kampaně, neboť preference a nákupní zvyklosti zákazníků jsou vysoce heterogenní. Znalost zákaznického profilu se tak stává klíčovou konkurenční výhodou (Hunt & Arnett, 2004). Jak ukazují moderní analýzy spotřebitelského chování (Trūata, 2021), pouhý ekonomický potenciál zákazníka již nutně neodráží jeho ochotu utracet. Kvantitativní metody, jako je shluková analýza (cluster analysis), proto dnes představují standardní nástroj pro tzv. tržní segmentaci. Umožňují odhalit skryté vzorce v datech a identifikovat specifické skupiny, jako jsou například impulzivní kupující či vysoce loajální zákazníci s velkým objemem útrat, na které lze následně cílit personalizované obchodní strategie (Kotler & Keller, 2016).

Pro efektivní cílení marketingových kampaní je klíčové porozumět struktuře zákazníků. V této zprávě se zabýváme segmentací zákazníků nákupního centra na základě jejich ročního příjmu a tzv. skóre útraty (Spending Score), které vyjadřuje míru jejich loajality a objem nákupů. Cílem je identifikovat typologické skupiny (shluky) zákazníků se specifickým nákupním chováním.

Pro tyto účely jsme využili veřejně dostupný datový soubor „Mall Customer Segmentation Data“ (Choudhary, 2018)¹, který obsahuje záznamy o 200 zákaznících. Pro naši analýzu jsme pracovali se dvěma klíčovými spojitými proměnnými: *Annual Income* (Roční příjem v tisících USD) a *Spending Score* (Skóre útraty na škále 1–100).

2. Metodologie a příprava dat

Data byla zpracována v programu TIBCO Statistica. Z důvodu rozdílných měřítek u obou proměnných (desítky až stovky tisíc u příjmu vs. jednotky až stovky u skóre) byla data před samotnou analýzou standardizována na z-skóre, aby se předešlo zkreslení výsledků vlivem odlišných variancí.

Pro rozdělení respondentů do skupin využijeme metod shlukové analýzy (cluster analysis). Pro odhad optimálního počtu skupin byla nejprve využita hierarchická shluková analýza. Princip této metody spočívá v tom, že se vypočítávají vzdálenosti mezi jednotlivými datovými body a na základě těchto vzdáleností se vytváří dendrogram. Rozhodli jsme se využít Wardovy metody, která je založená na minimalizaci součtu kvadratických odchylek v každém kroku spojování shluků. Výhodou této

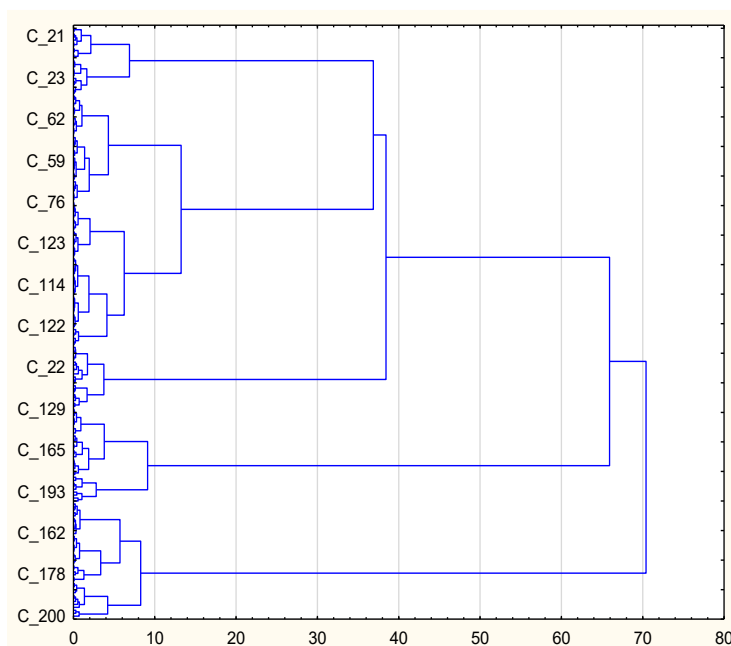
¹ Data a další informace o této zprávě jsou dostupné na adrese:
https://github.com/sharmaroshan/Clustering-of-Mall-Customers/blob/master/Mall_Customers.csv

metody je, že se snaží minimalizovat rozdíly mezi jednotlivými shluky a může vést k tvorbě poměrně homogenních shluků s co nejmenší možnou vnitřní variabilitou.

Následně byla aplikována metoda K-Means. Princip metody K-Means spočívá v tom, že se vybere počet K shluků, do kterých se budou data rozdělovat. V dalším kroku se náhodně vybere K centroidů, což jsou bodové reprezentace každého shluku. Poté se data přiřadí ke shlukům na základě nejmenší vzdálenosti k centroidu každého shluku.

3. Výsledky

Hierarchická shluková analýza (Odhad počtu shluků) je znázorněna na grafu 1. Na ose Y dendrogramu vidíme jednotlivá pozorování, která se postupně slučují do větších klastrů. Na ose X lze vidět spojovací vzdálenost mezi shluky spočítanou pomocí Wardovy metody. Z vizuální inspekce stromového grafu vyplývá, že při provedení pomyslného řezu na hladině spojovací vzdálenosti přibližně 10 se data přirozeně a stabilně dělí do pěti hlavních větví. Z tohoto důvodu bylo pro další krok zvoleno rozdělení do 5 shluků.

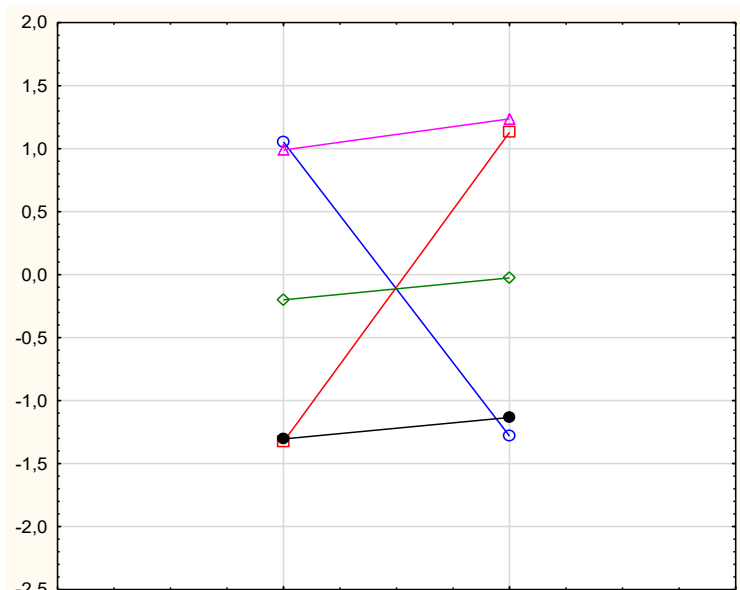


Graf 1:

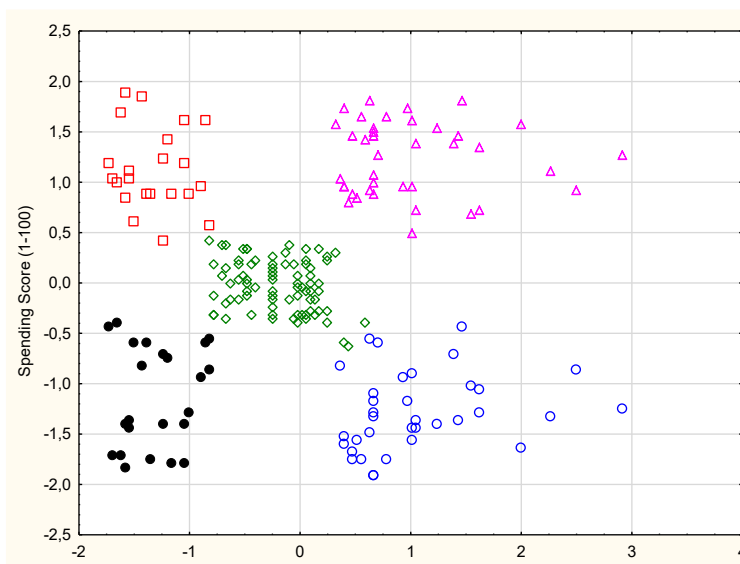
Stromový graf hierarchické shlukové analýzy (Wardova metoda)

Metoda K-Means (Profilování zákazníků)

Po stanovení počtu klastrů ($K=5$) byli zákazníci rozděleni do finálních skupin. Profil jednotlivých skupin ilustruje Graf průměrů (Plot of Means) standardizovaných hodnot a následný bodový graf (Scatterplot).



Graf 2: Graf průměrů pro jednotlivé klastry (standardizované hodnoty)



Graf 3:

Bodový graf zákazníků rozdělených do klastrů podle příjmu a útraty

Na základě grafického výstupu (Graf 2 a Graf 3) lze identifikovat pět výrazně odlišných zákaznických segmentů:

- **Klastr 1 (Šetřiví boháči / Modrá barva):** Zákazníci s nadprůměrným ročním příjmem, avšak s výrazně podprůměrným skóre útraty. Představují nevyužitý potenciál.
- **Klastr 2 (Impulzivní / Červená barva):** Zákazníci s podprůměrným ročním příjmem, kteří ale vykazují velmi vysoké skóre útraty.
- **Klastr 3 (Průměrní / Zelená barva):** Nejpočetnější a nejvíce nahloučená skupina ve středu grafu. Mají průměrné příjmy i průměrné nákupní chování.

- **Klastr 4 (VIP klienti / Fialová barva):** Zákazníci s vysokým příjmem a zároveň vysokým skóre útraty. Jedná se o nejcennější segment nákupního centra.
- **Klastr 5 (Skromní / Černá barva):** Zákazníci s nízkými příjmy a velmi nízkou ochotou utracet.

4. Závěr

Kombinací hierarchické metody a metody K-Means se podařilo úspěšně rozdělit 200 zákazníků do 5 homogenních segmentů. Nejdůležitějším zjištěním je identifikace „VIP“ skupiny (Klastr 4), na kterou by se měly zaměřit věrnostní programy, a skupiny „Šetřivých“ (Klastr 1), na kterou by bylo vhodné cílit specifické marketingové pobídky ke zvýšení jejich útraty, jelikož k tomu mají ekonomický potenciál.

Zdroje

- Choudhary, V. (2018). Mall Customer Segmentation Data. Kaggle. <https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python>
- Kotler, P., & Keller, K. L. (2013). Marketing management (14. vydání). Grada.
- Hunt, S., & Arnett D. (2004) Market Segmentation Strategy, Competitive Advantage, and Public Policy. *ResearchGate*. [https://doi.org/10.1016/S1441-3582\(04\)70083-X](https://doi.org/10.1016/S1441-3582(04)70083-X)
- TIBCO Software Inc. (2020). TIBCO Statistica (Verze 14.0) [Počítačový software].
- Truata. (2021). Global Consumer State of Mind Report 2021. <https://www.truata.com/wp-content/uploads/2026/01/Global-Consumer-State-of-Mind-Report-2021.pdf>