

Fakulta informačních technologií VUT v Brně

# Získávání znalostí z databáze transakcí studentského klubu

Zadání

Ondřej Ondryáš, Filip Čižmár

20. října 2023

# 1 Datová sada

Úlohy budou řešeny nad daty ze skutečného provozu studentského klubu U Kachničky na FIT VUT. Všichni zákazníci klubu se musí registrovat v informačním systému, při každém nákupu se prokazují členským průkazem a objednávky jsou spárovány s jejich profily, data tedy přesně zachycují jejich chování.

Databáze obsahuje data o transakcích v období cca červen 2021 až září 2023. Sledovanými typy transakcí jsou nákup, poskytnutí dobrovolného členského příspěvku, inventura zásob a naskladnění zásob. Dále jsou v ní reprezentovány související informace o uživateli, nabízených produktech a jejich štítcích. Některé entity typu produkt reprezentují neprodejné skladové položky, ze kterých se skládají jiné entity typu produkt. V databázi jsou proto oddělena data zachycující pohyby prodejních produktů a pohyby všech skladových položek. Schéma databáze je zobrazeno na diagramu v příloze A.

Data jsou uložena v relačním systému PostgreSQL, pro účely získávání znalostí tedy bude patrně nutné je extrahovat a transformovat do vhodné podoby, to by však mělo být otázkou několika SQL dotazů.

## 1.1 Anonymizace a úprava dat

Použitá databáze vznikla kopií produkční databáze dne 6. 10. 2023. Z kopie byly následně odstraněny některé relace a atributy, které v zamýšlených úlohách nemají význam. Odstraněny byly především informace o cenách a finančních tocích. Osobní údaje uživatelů byly nahrazeny charakteristikami, které neumožňují identifikaci konkrétní osoby.

# 2 Úlohy získávání znalostí

Cílem práce je vytvořit modely pro odhad spotřeby zásob a pro odhad počtu návštěvníků v klubu, a to za účelem optimalizace objednávek zásob a plánování směn obsluhy. Dalším cílem je rozdělit zákazníky do předem neznámých skupin podle chování při nakupování, a to za účelem identifikace typů zákazníků a vylepšování nabídky. Poslední úlohou je výpočet frekventovaných a sekvenčních vzorů za účelem vylepšení organizace uložení zásob v klubu.

## 2.1 Optimalizace objednávek zásob

Do klubu se pravidelně nakupují v zásadě tři kategorie zásob: výčepní nápoje (sudy); drobné občerstvení a suroviny pro přípravu toastů. Spotřeba všech tří se razantně mění v závislosti na postupu semestru: přirozeně zákazníků ubývá se zvyšováním studijní zátěže; tyto změny však nejsou lineární a spotřeba jednotlivých typů zásob se vyvíjí různým způsobem. Jednotlivé typy zásob navíc mají poněkud jiný charakter nákupu: sudy se objednávají cca jednou týdně, drobné občerstvení jednou za měsíc a suroviny bezprostředně před otevřením. V současnosti se zásoby nakupují čistě na základě intuice, což není optimální.

Je tedy žádoucí vytvořit regresní model, který na základě data (resp. kombinace čísla týdne semestru a dne v týdnu) a typu zásob určí vhodný objem k nakoupení.

## 2.2 Plánování směn obsluhy

Klub obsluhují dobrovolníci bez nároku na odměnu, typicky je otevřeno dvakrát týdně po dobu 6 hodin, obsluhu je tedy možné (a nutné) plánovat vcelku dynamicky, po hodinách.

Nutné personální zdroje jsou přitom závislé na počtu návštěvníků – na začátku semestru může být potřeba i dvakrát více obsluhujících než na jeho konci.

Počet návštěvníků je závislý na postupu semestru, ale značnou měrou i na termínu otevření (obvykle je otevřeno v pondělky a ve středy, kdy bývá návštěvnost větší). Pro plánování směn je nutné odhadovat návštěvnost po jednotlivých hodinách. Zde je tedy opět žádoucí vytvořit regresní model, který na základě data (čísla týdne semestru a dne v týdnu) a hodiny určí přibližný počet návštěvníků v klubu. Z dat bude nutné odfiltrovat speciální akce, zajímavé jsou zde pouze obvyklé otvírací hodiny.

Zajímavý by také mohl být model, který na základě zadaných atributů návštěvníka odhadne, jak dlouho se zdrží. Tato informace pro nás však není stěžejní.

### **2.3 Segmentizace zákazníků**

Nabídka v klubu je dosud také sestavována pouze intuitivně (zejm. v kategorii drobného občerstvení), přitom je zjevné, že je návštěvníky možné rozdělit do jistých skupin podle charakteru jejich nákupů – skladby objednávek (množství a typu produktů) a jejich frekvence. Zjištění takovýchto skupin a jejich zastoupení mezi návštěvníky nám pomůže určit, ve kterých typech produktů má smysl nabídku rozšiřovat, a které jsou naopak spíše okrajové.

### **2.4 Frekventované a sekvenční vzory**

Informace o konkrétních produktech (nebo úzkých množinách produktů), které návštěvníci nakupují společně, bychom využili k lepšímu uspořádání zásob v prostoru baru a ve skladech – společně zakupované položky by měly být blíže u sebe.

# Přílohy

## A ER diagram vstupní databáze

