

Téma:

Analýza dat z e-shopu o prodeji kosmetického zboží

Zdroj dat: <https://www.kaggle.com/mkechinov/ecommerce-events-history-in-cosmetics-shop>

Výstup: Odevzdáváte zip soubor – pojmenovaný: „MDAV_2_<vase_prijmeni>“

Rozsah: **2-6 stran textu** (do rozsahu textu se nepočítají tabulky a grafy, tj práce může mít i 20 stran, pokud budou výstupy opodstatněné!)

Dobrý den,

Vaším cílem bude souhrnná analýza dat a návrh systému měření kvality / prodeje pro více přiložený eshop zabývající se prodejem kosmetiky. Datový soubor obsahuje zhruba 20 milionu záznamů událostí za 5 měsíců.

Data jsou rozdělená do souborů po jednotlivých měsících, které jsou zhruba 0.5 GB velké. Je zcela na vás, jaký postup zpracování dat zvolíte, zda celé na jednou či po částech – jistě to bude limitováno i případnými výpočetní kapacitou! Data zpracujte samozřejmě v jazyku R – komplet!

Práce se zaměří na:

- Podrobnou evaluaci stávajícího stavu
 - o Budu chtít vědět, jak na tom daný e-shop je, co prodává a jaká je jeho situace. Popište daný e-shop a proveďte shrnutí situace.
 - o Samozřejmě musíte provést i analýzu stavu dat -> v jakém jsou stavu, jsou použitelná pro analýzu? Musíte odstranit nějaké záznamy?
- **Konstrukce a návrh indikátorů** evaluující to, jak se danému e-shopu daří.
 - o Budu chtít od vás teoretickou diskuzi nad tím jaké ukazatele by si měl e-shop hlídat
 - Tedy typ ukazatele, jak by se měl počítat (vzorec), v jaké periodě (agregace za jaký čas?) a co by daný ukazatel měl vysvětlovat.
 - o Budu chtít po vás Reálnou konstrukci všech relevantních ukazatelů pro daný e-shop -> Graficky vizualizovat.
 - o Dané indikátory musí poskytovat jak pohled na například hodinu provozu, tak ale budou i v denní, týdenní a měsíční verzi!
- Identifikace kritických situací v minulosti a jejich analýza
 - o Nastali v historii e-shopu nějaká období která naznačují to, že by docházelo k nějakým problémům?
 - Jak jste dané situace identifikovali? Proč jste je tak identifikovali? Postihují je vaše ukazatele?
 - o Zjistěte, v jakých kategoriích došlo, k jakým propadům, navrhnete systém a konkrétní pravidla která upozorní na případné problémy v budoucnu
 - Ukažte, jak daná pravidla by indikovala a zachovala se v daných 5 ti měsících.
 - Např: pokud dochází propadu ukazatele XY o více než 10 % tak nastává kritická situace – a ukázat minulé kritické situace a budu chtít vysvětlit proč k nim došlo a zda je o opravdu vaše pravidla pro dané ukazatele identifikovala.

- Výhled do budoucna:
 - Jaké další data by bylo vhodné za daný e-shop sbírat?
 - Jaké další změny byste navrhovali?

Poznámky:

- ➔ Výstupy by měli být prezentované formou grafů a tabulek v zahrnuté do seminární práce.
- ➔ K seminární práci přiložíte kód který bude dokumentovat celý váš postup -> každá tabulka, číslo i graf (!) musí být replikovatelný a musím spuštěním daného kódu ověřit to jak jste k daným hodnotám /grafům došli!
- ➔ Je zcela na vás, pokud si daná data agregujete či vytvoříte dílčí soubory, jejich postup vytvoření a zpracování ale musí být součástí kódu.
- ➔ Součástí kódu bude krátká dokumentace (formou komentářů) dokumentují váš postup v dané části kódu.
 - Netestuji (tedy nehodnotím) to, zda umíte programovat, ale testuji to, zda váš postup byl správný (z pohledu konstrukce daných ukazatelů a práce s daty) a musím to být schopen pochopit – co jste dělali a proč!
- ➔ **Text práce má mít délku opravdu od 2 do 6 ti stran. Snažte se psát výstižně a stručně ale tak abych pochopil co jste tím mysleli!**
- ➔ **Grafické výstupy / tabulky se do toho textu opravdu nepočítají !!! objem grafických výstupů vložených do textu je zcela na vás, ale měl by vystihovat a vysvětlovat všechny diskutované problémy!**
- ➔ Komentáře a dotazy k této práci prosím piště pod vložený soubor do skupiny na teamsech.