

Výuková data Hotel

Popis dat

verze 11. 2. 2015

Vypracoval: M. Šimůnek

Data k analýze

Sloučená data:

- <http://lispminer.vse.cz/data/Hotel/HotelPlusExterni.1000.txt>

Počet záznamů: 1000

Období: 2012 až 2013

Dílní soubory (pro odzkoušení připojování externě dodaných dat):

- pobyty hostů: <http://lispminer.vse.cz/data/Hotel/Hotel.1000.txt>
- počasí: <http://lispminer.vse.cz/data/Hotel/Hotel.Weather.txt>
- směnný kurz CZK/EUR: <http://lispminer.vse.cz/data/Hotel/Hotel.Exchange.txt>

Soubory s počasím a směnným kurzem obsahují jeden záznam pro každý den z období 2012 až 2013, tedy 731 záznamů.

Základní informace

Databáze pobytů hostů v nejmenovaném hotelu. Pro účely výuky byl soubor omezen na 1000 záznamů. Základní data byla vygenerována v aplikaci *LM ReverseMiner*. Kromě údajů o pobytech byly do datové sady přidány externí údaje o průměrné denní teplotě (z webové stránky <http://www.isws.illinois.edu>) a dále údaje o směnném kurzu CZK/EUR z webové stránky ČNB.

Kvalita je v případě dat o hotelu neznámá. Při analýze je třeba brát v úvahu možné zkreslení předzpracování i výsledků vlivem špatně zadaných hodnot. V případě meteorologických měření byl pro každý sloupec uveden kód správnosti hodnoty (nebo jejího odhadu). Vzhledem k naprostému minimu dopočítávaných hodnot a pro účely výuky je možné tento rozdíl zanedbat a všechny hodnoty počasí považovat za správné. Stejně tak jsou spolehlivé i hodnoty směnných kurzů.

Popis sloupců

Jako primární klíč je možné použít při importu uměle přidáný sloupec *ID_LM*.

Hotel

Zdroj: LM ReverseMiner

Sloupce uvozeny písmenem 'H'.

Skupina	Sloupec	Datový typ	Popis
Guest	HAge	Integer	věk hlavního hosta
Guest	HNationality	Text	národnost (kódy ISO)
Guest	HPersons	Integer	počet osob (včetně hlavního hosta)
Visit/Calendar	HVisitFrom	Date	den příjezdu
Visit/Calendar	HNights	Integer	počet nocí
Visit	HPrice	Integer	celková cena v Kč
Visit	HRoom	Text	číslo (název) pokoje
Visit	HVisitType	Text	typ pobytu (turista, služební cesta)

Počasí

Zdroj: LM ReverseMiner, <http://www.isws.illinois.edu/warm/data/cdfs/stcday.txt>

Sloupce uvozeny písmenem 'W'.

Skupina	Sloupec	Datový typ	Popis
Weather	WSky	Text	slunečno, zamračeno, déšť/sněžení
Weather	WTempAvg	Decimal	teplota vzduchu, průměrná [°C]

V samostatném souboru <http://lispminer.vse.cz/data/Hotel/Hotel.Weather.txt> jsou navíc další sloupce (převzaté z <http://www.isws.illinois.edu/warm/data/cdfs/stcday.txt> a převedené na metrické jednotky):

Skupina	Sloupec	Datový typ	Popis
		typ	
Weather	wwind_speed_gust_kmh	Decimal	rychlost větru, nárazová [km/h]
Weather	wwind_speed_avg_kmh	Decimal	rychlost větru, průměrná [km/h]
Weather	wwind_dir_deg	Decimal	směr větru, azimut [stupně]
Weather	wsolar_rad_MJ	Decimal	sluneční záření [megajouly]
Weather	wtemp_max_C	Decimal	teplota vzduchu, maximální [°C]
Weather	wtemp_min_C	Decimal	teplota vzduchu, minimální [°C]
Weather	wtemp_avg_C	Decimal	teplota vzduchu, průměrná [°C]
Weather	wprecip_mm	Decimal	množství srážek, vodní sloupec [mm]

Hodnoty *WSky* byly generovány náhodně, a proto nemají žádný vztah k *wsolar_rad_MJ* a *wprecip_mm*.

Směnné kurzy

Zdroj: http://www.cnb.cz/cs/financni_trhy/devizovy_trh/kurzy_devizoveho_trhu

Záznamy byly doplněny, aby obsahovaly kurz pro každý den (včetně víkendů a svátků).

Sloupce uvozeny písmenem 'X'.

Skupina	Sloupec	Datový typ	Popis
Exchange	XEUR	Decimal	směnný kurz CZK/EUR v den příjezdu

Odvozené hodnoty

Ze základních hodnot lze dopočítat odvozené hodnoty. Kromě hodnot odvozených z datového typu Datum (den v týdnu, čtvrtletí...) je možné přidat například tyto:

Skupina	Sloupec	Datový typ	Popis
Visit	HPriceBase	Float	základní cena za osobu a noc vypočtena jako: $HPrice / (HNights * HPersons)$
Exchange	XPriceBaseEUR	Float	základní cena v eurech vypočtena jako: $HPrice / (XEUR * HNights * HPersons)$
Visit/Calendar	HSaturdayNight	Ano/Ne	pobyt trval alespoň přes jednu sobotní noc. Vypočteno výrazem pomocí jazyka VBA jako: $Weekday(HVisitFrom, 1) + HNights >= 8$