Tato prezentace je součástí wiki-prezentace Metoda GUHA, LISp-Miner a typové úlohy
Je dostupná z této adresy
Verse 14. 8. 2019

Typ úlohy: Klesající histogram
Data: Hotel
Problém: Existuje segment hostů s rostoucím trendem spokojenosti, tedy % nespokojeno > % průměrné hodnocení > % spokojeno?

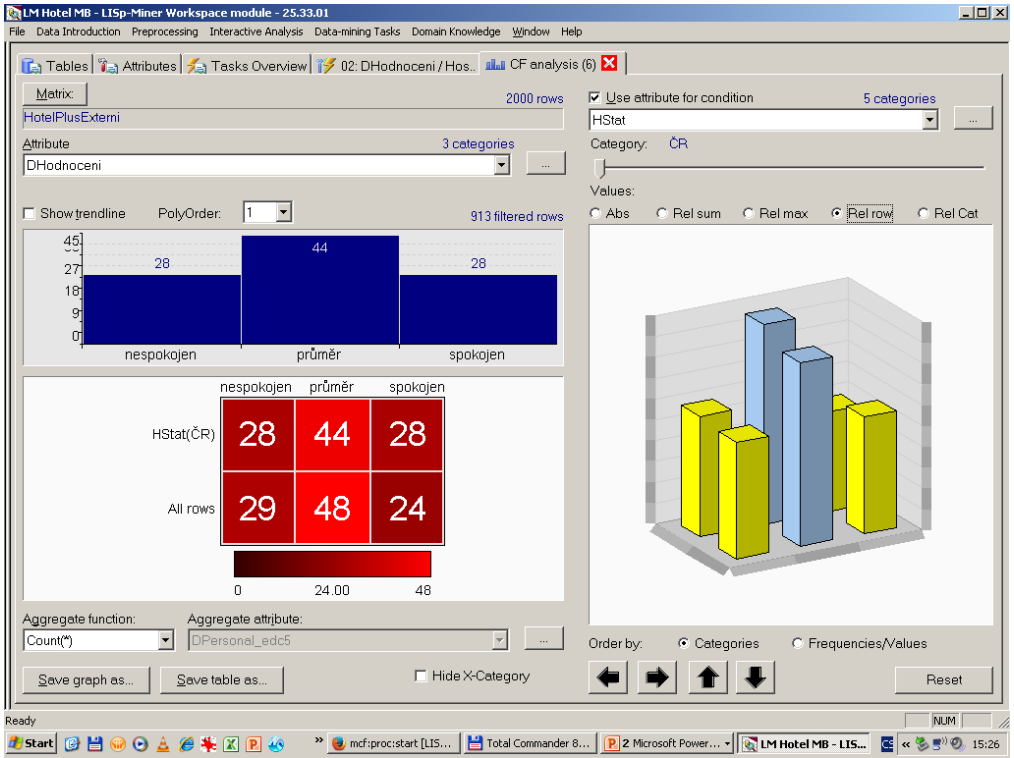
Jan Rauch
Katedra informačního a znalostního inženýrství
Vysoká škola ekonomická v Praze

© Jan Rauch

Histogram klesá

Inspirace - Hosté z ČR:

28% nespokojeno, 44% průměrné hodnocení, 28% spokojeno



Existuje segment hostů s klesajícím trendem spokojenosti, tedy
% nespokojeno > % průměrné hodnocení > % spokojeno?
Zajímá nás Pobyt, Host, Bydliště, Meteo

Aplikace procedury CF-Miner – histogram klesá

Hotel.txt

HVek	HPohlavi	HMesto	HMesto_X	HMesto_Y	HStat	PPobytOd	PNoci	POsob	PTypPobytu	PCenaUbytovani	PCenaStrava	PCenaSleva	PCenaCelkem	DHodnoceni			
21	žena	České Budějovice	14.4757883	48.9763169	ČR	31.5.2013	1	1	rekreační	1450.00	0.000	0.00	1450.00	spokojen	91	82	81	56
34	muž	Linec	14.2862742	48.3066489	Rakousko	2.8.2013	2	4	rekreační	11600.00	1440.000	200.00	12840.00	průměr	44	21	62	84
30				48.211563														
62			1283	50.25														
35			767	48.2115														
58	muž	České Budějovice	14.475788															
81	žena	Vídeň	16.3736767	48.2115														
22	žena	Drážďany	13.7397044	51.														
82	muž	Katovice	19.0241283	50.25														
55	muž	Praha	14.4212806	50.087496														
75	žena	Berlín	13.3908886	52.5176189	Německo	26.5.2012	1	4	rekreační	4840.00	0.000	200.00	4640.00	nespokojen	30	39	32	3
66	žena	Linec	14.2862742	48.3066489	Rakousko	23.2.2012	1	2	služební	2420.00	0.000	0.00	2420.00	spokojen	95	93	75	21
64	žena	Linec	14.2862742	48.3066489	Rakousko	16.3.2013	7	2	rekreační	20300.00	0.000	600.00	19700.00	průměr	52	30	59	52
35	muž	Košice	21.2543528	48.7160408	Sl													
32	muž	Mnichov	11.5836375	48.1364669	Ně													
65	muž	Plzeň	13.3771556	49.7490406	ČR													
79	muž	Brno	16.6153758	49.1921808	ČR	3.5.2012	1	3	rekreační	3630.00	0.000	0.00	3630.00	průměr	59	43	48	59
28	žena	Drážďany	13.7397044	51.0497456	Německo	24.12.2012	1	2	služební	2420.00	0.000	0.00	2420.00	průměr	51	66	58	48
35	žena	Hamburg	10.0043528	53.5498325	Německo	5.1.2013	14	2	rekreační	40600.00	5040.000	200.00	45440.00	nespokojen	27	15	30	12
22	muž	Plzeň	13.3771556	49.7490406	ČR	9.11.2013	4	1	rekreační	5800.00	0.000	200.00	5600.00	spokojen	84	79	86	48
25	muž	Karlovy Vary	12.8690381	50.2311075	ČR	9.11.2013	7	2	rekreační	20300.00	0.000	600.00	19700.00	spokojen	88	74	94	54
20	žena	Hamburg	10.0043528	53.5498325	Německo	19.1.2013	14	4	rekreační	81200.00	10080.000	200.00	91080.00	průměr	59	23	46	96
30	žena	Linec				29.11.2013	2	4	rekreační	11600.00	0.000	200.00	11400.00	spokojen	88	82	84	61
45	žena	Karlov			R	28.12.2012	2	2	rekreační	4840.00	600.000	200.00	5240.00	spokojen	81	99	87	23

Existuje segment hostů s klesajícím trendem spokojenosti, tedy % nespokojeno > % průměrné hodnocení > % spokojeno?
Zajímá nás Pobyt, Host, Bydliště, Meteo

? CF: StDn $_{2,\ \geq 100}$ DHodnocení / Host, Bydliště, Meteo, Pobyt

Meteo.txt

MDatum	MTeplota	MObloha
4.1.2012	-6.3	slunečno
5.1.2012	-6.6	zataženo
6.1.2012	6.1	srážky
7.1.2012	1.6	srážky
8.1.2012	-1.3	srážky
	7	zataženo
		srážky
		zataženo
12.1.2012	-3.1	srážky
13.1.2012	-8.1	zataženo
14.1.2012	-10.7	srážky
15.1.2012	-5.5	zataženo
16.1.2012	2.3	zataženo
17.1.2012	-1.9	zataženo
18.1.2012	-8.6	zataženo

? CF: StDn $_{2,\ \geq 100}$ DHodnocení / Host, Bydliště, Meteo, Pobyt, zadání pro CF-Miner

LM Hotel_Toshiba MB - LISp-Miner Workspace module - 25.37.00

File Data Introduction Preprocessing Interactive Analysis Data-mining Tasks Domain Knowledge Window Help

Tables | Attributes | Tasks Overview | KL analysis (5) | 03: PDenTydne / Poby.. | 03: PDenTydne / Poby.. | 03: PDenTydne / Poby.. | 03: PDenTydne / Poby.. | Hypothesis

Data-mining Task basic parameters

Name: 03: PDenTydne / Pobyt, Host, Bydliště, Meteo StDn

ID: 14

Comment: -

Taskgroup: 01: CF-Miner

Task type: CF-Miner

Data matrix: HotelPlusExterni

Edit

ATTRIBUTES FOR HISTOGRAM

Attributes

» DHodnoceni(3 categories)

DHodnocení

QUANTIFIERS

Type	Source	Range	Rel.	Value	Units
SUM	Abs	all	>=	100.00	Abs
S-DN	%Cond	all	>=	2.00	Abs

StDn $_{2,\ \geq 10}$

CONDITION

Pobyt	Con, 1 - 4
» PNoci_enum_m (seq), 1 - 2	B, pos
» PDenTydne (subset), 1 - 1	B, pos
» POsob (seq), 1 - 3	B, pos
» POsobonoci_ef5 (seq), 1 - 2	B, pos
Host	Con, 0 - 2
» HPohlavi (subset), 1 - 1	B, pos
» HVek_exp (seq), 1 - 2	B, pos
Bydliště	Con, 0 - 1
» HCizinec_b (subset), 1 - 1	B, pos
» HMesto (subset), 1 - 1	B, pos
» HStat (subset), 1 - 1	B, pos
Meteo	Con, 0 - 2
» MObloha (subset), 1 - 1	B, pos
» MTeplota_exp (seq), 1 - 2	B, pos

Generation information

Status: Solved, 16 run(s)

Mode: Standard

Aggregate function

Type: Count(*)

Attribute: -

Task parameters

Include 'worse' extensions of condition: Yes

Include extensions of coefficients with no change in the histogram: Yes

Include extensions of cedents with no change in the histogram: Yes

Maximal number of hypotheses:

Extensions minimal length check:

Pobyt (*) $\wedge$ Host(*) $\wedge$
$\wedge$ Bydliště(*) $\wedge$ Meteo(*)
Viz příklad 2

Params | Validate | Task Clone

Run | Bkgrnd Run | Grid Run | Show Results

Jedno z mnoha možných zadání pro řešení dané analytické otázky.

Ready

NUM

CF-kvantifikátor SUM

QUANTIFIERS

Type	Source	Range	Rel.	Value	Units
SUM	Abs	all	>=	100.00	Abs
S-DN	%Cond	all	>=	2.00	Abs

CF Simple frequencies quantifier settings

Interest measure type: Sum of frequencies

Sum of frequencies from range of categories

Relation: Greater than or equal

Threshold value: 100

Source frequencies: Absolute number

Absolute frequencies. Frequencies in histogram are left as given.

Threshold-value units: Absolute number

Absolute number of rows. Threshold value is left as given.

Category Range

From: 0 To: 100

Reset to All

Primary IM Settings

SUM >= 100 minimálně počet pobytů v segmentu

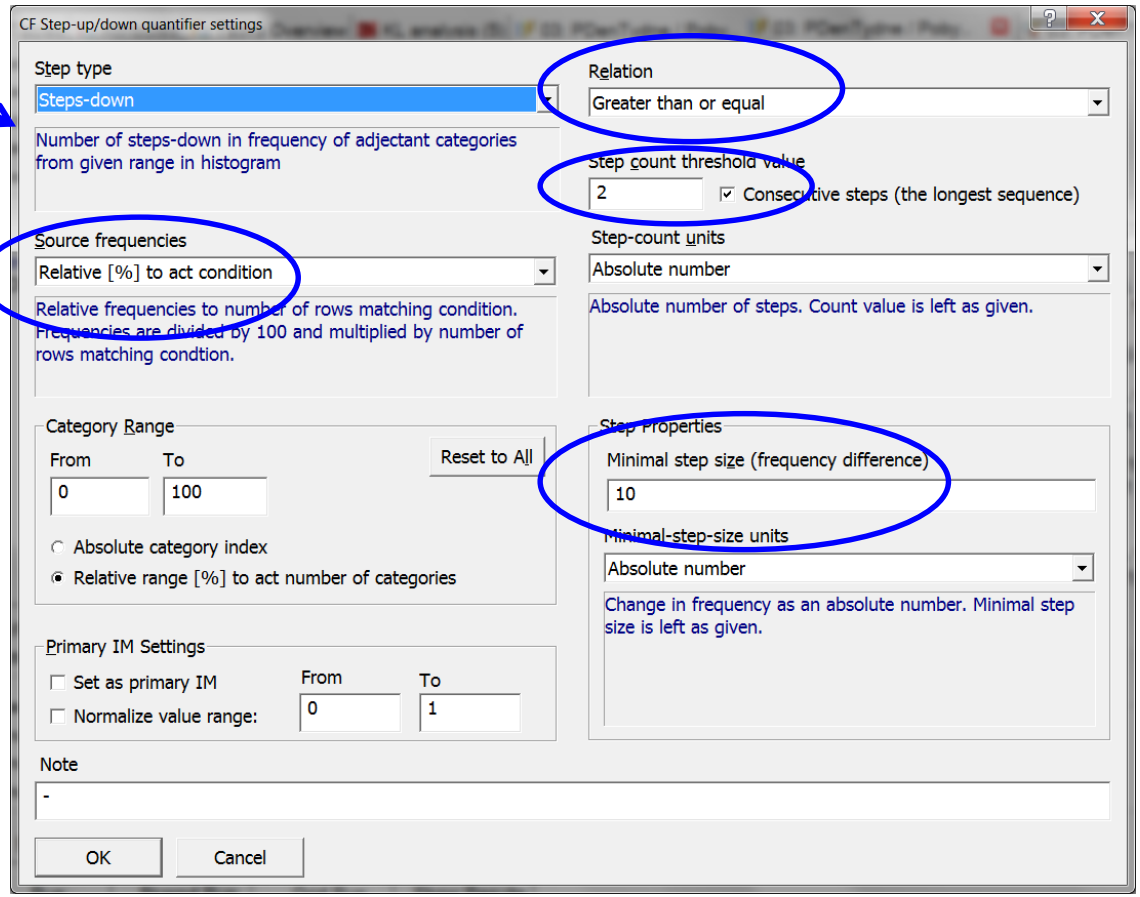
Note: -

OK Cancel

CF-kvantifikátor StepsDown

QUANTIFIERS

Type	Source	Range	Rel.	Value	Units
SUM	Abs	all	>=	100.00	Abs
S-DN	%Cond	all	>=	2.00	Abs



? CF: StDn $_{2,\ \geq 100}$ DHodnocení / Host, Bydliště, Meteo, Pobyt, výstup procedury CF-Miner

Task run

Start: 5.4.2016 09:54:44 Total time: 0h 0m 1s

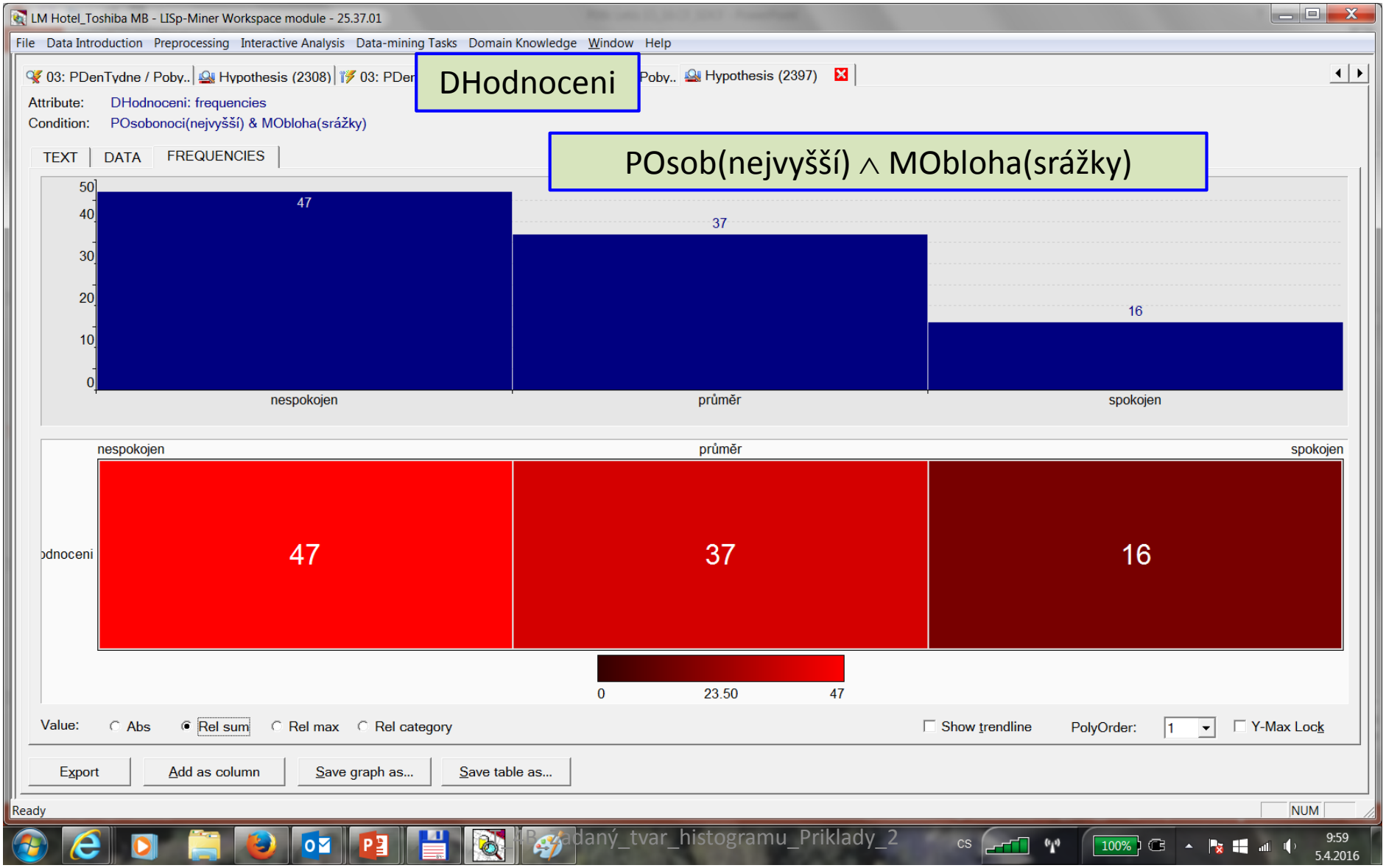
Number of verifications: 24935

Number of hypotheses: 36 Mode: Standard

Add group | Del group | Edit group

Actual group of hypotheses: All hypotheses

Hypotheses in group: 36 Shown hypotheses: 36 Highlighted: 0

Nr.	Id	Sum	Hypothesis
1	6	160	**DHodnoceni** / **PNoci**(*2*) & **PDenTydne**(*Pá*) & **MObloha**(*slunečno*) & **MTeplota**(*zima*)
2	5	160	**DHodnoceni** / **PNoci**(*2*) & **PDenTydne**(*Pá*) & **MObloha**(*slunečno*) & **MTeplota**(<=*zima*)
3	12	160	**DHodnoceni** / **PNoci**(*2,<3;6>*) & **PDenTydne**(*Pá*) & **MObloha**(*slunečno*) & **MTeplota**(*zima*)
4	11	160	**DHodnoceni** / **PNoci**(*2,<3;6>*) & **PDenTydne**(*Pá*) & **MObloha**(*slunečno*) & **MTeplota**(<=*zima*)
5	35	139	**DHodnoceni** / **POsobonoci**(*nejvyšší*) & **MObloha**(*srážky*)
6	3	129	**DHodnoceni** / **PNoci**(*2*) & **HCizinec**(*ne*) & **MTeplota**(<=*zima*)
7	4	129	**DHodnoceni** / **PNoci**(*2*) & **HStat**(*ČR*) & **MTeplota**(<=*zima*)
8	36	121	**DHodnoceni** / **POsobonoci**(*nejvyšší*) & **HVek**(>=*od 28 do 60*) & **MObloha**(*srážky*)
9	30	115	**DHodnoceni** / **PNoci**(*21,<22;27>*) & **POsobonoci**(*nejvyšší*) & **MTeplota**(*neutrální,teplo*)
10	28	115	**DHodnoceni** / **PNoci**(*21,<22;27>*) & **POsobonoci**(>=*vyšší*) & **MTeplota**(*neutrální,teplo*)
11	23	115	**DHodnoceni** / **PNoci**(*21,<22;27>*) & **MTeplota**(*neutrální,teplo*)
12	19	114	**DHodnoceni** / **PNoci**(*7*) & **MTeplota**(*teplo*)
13	20	114	**DHodnoceni** / **PNoci**(*7*) & **MTeplota**(>=*teplo*)
14	17	114	**DHodnoceni** / **PNoci**(*<3;6>,7*) & **MTeplota**(*teplo*)
15	18	114	**DHodnoceni** / **PNoci**(*<3;6>,7*) & **MTeplota**(>=*teplo*)
16	21	114	**DHodnoceni** / **PNoci**(*7,<8;13>*) & **MTeplota**(*teplo*)

Detailní výstup pátého histogramu