Практична робота №4 Використання інструментів ''HighlightExceptions'' і ''ScenarioAnalysis''

анотація: Практична робота присвячена використанню інструментів "Виділення винятків" ("HighlightExceptions") і "Аналіз сценаріїв" ("ScenarioAnalysis").

В якості навчального набору даних, як і в минулих лабораторних, будемо використовувати локалізований приклад для Excel, взятий з http://russiandmaddins.codeplex.com/

Виділення винятків

Як випливає з назви, інструмент дозволяє виявити дані, що виділяються серед наявного набору. Це може бути корисно в ряді випадків. По-перше, це можуть бути помилкові дані (наприклад, результати помилки оператора при введенні якихось значень). По-друге, виключення можуть становити окремий інтерес (як, наприклад, в разі виявлення шахрайських дій з банківськими картами і т.п.). Крім того, аналіз винятків може розглядатися як попередня частина інтелектуального аналізу даних за допомогою інших методів. Зокрема, це дозволяє виключити потрапляння нетипових прикладів в навчальну вибірку.

В ході роботи інструмент HighlightExceptions створює тимчасову модель інтелектуального аналізу з використанням алгоритму MicrosoftClustering. Для кожної аналізованої рядки оцінюється ступінь приналежності виявлених кластерам.Значенія, що знаходяться далеко від всіх кластерів, позначаються як виключення.

При запуску інструменту можна відзначити стовпці, що не враховуються при аналізі. У рекомендаціях вказується, що бажано виключити з аналізу стовпці з унікальними значеннями (імена, ідентифікатори), а також містять багато порожніх значень або довільний текст. На Рис. 4.1видно, що при аналізі набору даних "Клієнти" інструмент пропонує виключити з розгляду поле ID. за підсумками роботи (а працює цей інструмент трохи довше розглянутих нами раніше) формується звіт (Рис. 4.2) І в початковому наборі даних виключення виділяються кольором (Рис. 4.3).

x	10	• (* · =	codepl	ex050311	_PI_Data_Mining - Min	rosoft	Excel	-		Patients	стаблицами		×
0	uin n	павная Бставка Разми	тка страни Формул	ы Дан	ные Рецензирования	Бид	Data Minir	ng Pa6o	чая группа	Analyze	Конструктор	a 🕜 🗆 🖨	13
Ana	lyze Key uencers	Detect Fill From Fore Categories Example	cast Highlight Scel Exceptions Anal	anio Pri tysis * Ca	ediction Shopping Iculator Basket Analysis	DM (lo	AddinsD8 (calhost)	Help					
	G	i7 • (*	<i>f</i> x Професс	ионал	SQL Server Data M	lining	- Highlight	Exceptio	ns			8 - X	
1	A	В	С	D	Highlight E	read	iner				Ton More	witz	
1	Sample	data for Analyze Key Inf	luencers, Detect Cat	egories	ringungut C	Acepa	AND .				SC	L Server 20	08
234	Прил	мер данных д/ Семейное положен	ия "Анализа ине 🕶 Пол 💽	влия Дохол	The Highlight Exceptio govern the column val The rows containing et	ns tool ues of n coeption	detects rows nost other ro ns are highlig	s in your to ws. The ti phted and	ible that do no col generates the column va	t conform t a detailed o slues that a	o the rules and g exceptions report re most likely to o	eneralizations that in a separate sheet cause the exception	t.
5	12496	Женатый, замужняя	Женский	40	are emphasized.								
6	24107	Женатый, замужняя	Мужской	30									
7	14177	Женатый, замужняя	Мужской	80	Column Selection								
8	24381	Одинокий(ая)	Мужской	70	Calact the caluma		unatte an	aluma fa					
9	25597	Одинокий(ая)	Мужской	30	Select the column	is you	want to an	alyze ru	exception	•			
10	13507	Женатый, замужняя	Женский	10	ID ID							<u> </u>	
11	27974	Одинокий(ая)	Мужской	160	Поп	кение							
12	19364	Женатый, замужняя	Мужской	40	И Доход								
13	22155	Женатый, замужняя	Мужской	20	Дети								
14	19280	Женатый, замужняя	Мужской	20	 Гип работы 							1	
15	22173	Женатый, замужняя	Женский	30	И Домовладелец								
16	12697	Одинокий(ая)	Женский	90	Pacctogere cou	afor	8						
17	11434	Женатый, замужняя	Мужской	170	Регион								
18	25323	Женатый, замужняя	Мужской	40	🕼 Возраст							-	
19	23542	Одинокий(ая)	Мужской	60									
20	20870	Одинокий(ая)	Женский	10									
For	080	Клиенты / Заполне	ние из примера	Прогн		_		_	_		Run	Close]

Рис. 4.1. Запуск інструменту HighlightExceptions

Собил С D E F G H 1 Highlight Exceptions Report for Taблица3 0 6 </th <th>-Circles</th> <th></th> <th></th> <th></th> <th></th> <th>rosoft Excel</th> <th>Mining - Mic</th> <th>1_PI_Dat</th> <th>plex05031</th> <th>codep</th> <th></th> <th></th> <th>• 🛛</th> <th>9.9.6</th> <th>X</th>	-Circles					rosoft Excel	Mining - Mic	1_PI_Dat	plex05031	codep			• 🛛	9.9.6	X
A Calibri 11 Image: Calibri 11 Image: Calibri 11 Image: Calibri Image: Calibri <thimage: calibri<="" th=""> <thimage: calibri<="" th=""> <thi< th=""><th>pynna 🛛 🕜 😅 i</th><th>чая группа б</th><th>Рабочая</th><th>ata Mining Pa</th><th>g Dat</th><th>ание Вид</th><th>Рецензиров</th><th>Данные</th><th>лулы ,</th><th>страницы Форм</th><th>Разметк</th><th>Вставка Рі</th><th>13.9</th><th>л Главн</th><th>0</th></thi<></thimage:></thimage:>	pynna 🛛 🕜 😅 i	чая группа б	Рабочая	ata Mining Pa	g Dat	ание Вид	Рецензиров	Данные	лулы ,	страницы Форм	Разметк	Вставка Рі	13.9	л Главн	0
Threshold fx 75 A B C D E F G H 1 Highlight Exceptions Report for Taблицa3 3 6 7 <	АТ Райти Сортировка Найти и фильтр * выделит Редактирование	∑ • ЯГ Сортиров 2• и фильтр Редактир	· Σ · Q	й ^{ча} Вставить * ₩ Удалить * Щ Формат * Ячейки	ание * лицу *	орматировани вать как таблиі к т	👫 Условное ф 😸 Форматирс 🕞 Стили ячее	- % 000	Общий 	= = = = = = = = = = 注注 ②・ Bupachineanine 5	11 · A A	н - 11 К Ч - А Ф - <u>А</u> Шрифт	Calif X E	анть 🛷	Вст
A B C D E F G H 1 Highlight Exceptions Report for Taблицa3	91610 310	V(0,0,)								75	fx	•	ł	Threshold	
Image: Provide and the state of t	1 1 2	H	1	G	F	E	D	C	в			А			A.
2 The outlier cells are highlighted in the original table. 3 Exception threshold (more or fewer exceptions) 7 Column • Outliers • 7 Cemeйное положение 0 8 10 Дети 11 Образование 12 Тип работы 13 Домовладелец 14 Кол-во авто 15 Расстояние до работы 16 Регион 17 Возраст						a3	or Таблиц	ort fo	ns Rep	ght Exception	Highl	н			1
3 4 Exception threshold (more or fewer exceptions) 7 Column • Outliers 7 Cemeйное положение 0 8 Пол 9 Доход 4 9 10 Дети 9 1 10 Бразование 11 1 12 Тип работы 13 Домовладелец 14 Кол-во авто 7 Регион 17 Возраст										ne original table.	ted in t	re highlighte	cells a	he outlier	2
4 Exception threshold (more or fewer exceptions) 75 5 ••••••••••••••••••••••••••••••••••••										in on prior to ore		0.0			3
4 Exception threshold (more or fewer exceptions) 75 5								-							
3 Column Image: Outliers								1	75	exceptions)	r fewer	ild (more or	hresh	exception th	4
о солони Облез С 7 Семейное положение 0 8 Пол 0 9 Доход 4 10 Дети 9 11 Образование 1 12 Тип работы 3 13 Домовладелец 0 14 Кол-во авто 7 15 Расстояние до работы 4 16 Регион 0 17 Возраст 6								1	all our la					a human	5
7 Семенное положение 5 8 Пол 0 9 Доход 4 10 Дети 9 11 Образование 1 12 Тип работы 3 13 Домовладелец 0 14 Кол-во авто 7 15 Расстояние до работы 4 16 Регион 0 17 Возраст 6								-	Ulers -			0000	0000	orumn Ceweğune r	0
9 Доход 4 10 Дети 9 11 Образование 1 12 Тип работы 3 13 Домовладелец 0 14 Кол-во авто 7 15 Расстояние до работы 4 16 Регион 0 17 Возраст 6									0			CHNC.	10/10/	Семенноет	8
О Дети 9 10 Дети 9 11 Образование 1 12 Тип работы 3 13 Домовладелец 0 14 Кол-во авто 7 15 Расстояние до работы 4 16 Регион 0 17 Возраст 6								1	4					Тохол	9
Образование 1 1 Образование 1 12 Тип работы 3 13 Домовладелец 0 14 Кол-во авто 7 15 Расстояние до работы 4 16 Регион 0 17 Возраст 6									ę					Іети	10
12 Тип работы 3 13 Домовладелец 0 14 Кол-во авто 7 15 Расстояние до работы 4 16 Регион 0 17 Возраст 6								1	1				ие	бразовани	11
13 Домовладелец 0 14 Кол-во авто 7 15 Расстояние до работы 4 16 Регион 0 17 Возраст 6									3					ип работы	12
14 Кол-во авто 7 15 Расстояние до работы 4 16 Регион 0 17 Возраст 6								í.	0				елец	омовладе	13
15 Расстояние до работы 4 16 Регион 0 17 Возраст 6								1	7				0	ол-во авто	14
16 Регион 0 17 Возраст 6								i	4			боты	до ра	асстояние	15
17 Bospacr 6								<u>k</u>	C					егион	16
								i	6					возраст	17
18 Приобрел велосипед 0								1	0			пед	велос	Іриобрел в	18
19 <u>Итог 34</u>									34					TOF	19
20 A 4 5 M Youngury Taforena Quitfore Zaponaujuo ya pouluona Doppidamonanua (21 D 4 m			1.00	DAG		0 01	0001030000000	0 / 0	2 DRUMOI	e monunum u	Outline	Tafamura? (autra a	A MI Vour	20
Trans Contents Lawrences Satometerie is reported a Technological and the Technological a	0 0	00% (=)	UII 1003	1000 1718 CON		COLOR -	рогнозгрован	10 2 1	тример	 Sanotheenne sti 	outlier	raomitjas t	04150	ROUND ROUND	Ent



Dula Inter	as Revalues									
100 A		Разнетка страницы Форн	рлы Данные Ре	Unicopositive	Eng	Data Mining Patio-sas rpyrns	Analyz	конструктор	0 0	00
nalyze Key Deta Muencers Catlep	ed Fill From P ories Example	Table Analysis Tools	tion Shopping Di ator Basket Analysis	MAddiniDB Received scorection						
J74	• (*)	£ 0-1 KM								
I flon +	Доход - Дет	пи • Образование •	Тип работы	Домовладеле	4 -	Кол-во авто - Расстояние до работ	Tes +	Penvou - Bospa	кт - Приобрел велосипед	
0 Мужской	30000	2 Неоконченное высшее	Офисный работник	Дa	01	Сортировка от д до Я		Россия	68 Her	-
1 женский	40000	0 Bucuee	Офисный работник	Dia.	=1	Содтировка от # до А.	1	Eepona	37 Дa	
2 Мунской	30000	0 Среднее	Ручной труд	Да	10	Coonsponse no users	-	Concernation	and a state of the	1
3 женский	20000	4 Среднее	Ручной труд	Да				coproposida no da	ery avenua	41
4 женский	10000	0 Неоконченное среднее	Ручной труд	Нет		Weinele density is anothered by harden	1	1		1
5 Мужской	120000	0 Неоконченное среднее	Профессионал	Да		Severa to meets				
5 женский	10000	0 Неоконченное среднее	Ручной труд	нет		Текстовые фильтры		Het Shandkal	in the second second second second	1
7 женский	130000	3 Среднее	Профессионал	Да		Понск	Q,	Пользователя	ская сортировка	
8 женский	20000	0 Несконченное высшее	Ручной труд	Нет		(Begenera ace)		Espona	36 Да	1
9 Женский	20000	3 Среднее	Квалифицированны	47 Het		20-149	1	Россия	62 Her	
0 женский	130000	4 Среднее	Управление	Да		S 10+ m	1	Россия	31 Her	
1 Женсний	20000	0 Неоконченное среднее	Ручной труд	Нет		12 1-2 x04	-8	Espona	26 Her	-21
2 Мунской	80000	0 Бакалаер	Профессионал	Да		£ 5-10 KH	1	Россия	29 Да	5
3 Мунской	80000	2 Среднее	Квалифицированны	й Нет				Россия	50 Да	
Мунской	40000	2 Бакалавр	Управление	Да			- 6	Россия	63 Да	
5 Женский	30000	4 Высшее	Офисный работник	Да				Eepona	45 Да	
6 женский	10000	4 Неоконченное среднее	Ручной труд	Да			_	Espona	40 Her	16
7 Мунской	30000	0 Бакалаер	Офисный работник	Да				Espona	47 Да	1
8 Мунской	20000	0 Среднее	Ручной труд	нет		OK OTHER	8.ii) [Espona	29 Her	
9 Мунской	40000	2.Бакалаер	Управление	Her	-	1 3-10 KM	- 5	Россия	52 Да	
0 Мужской	10000	0 Неоконченное высшее	Ручной труд	Да		1-1-2 xxx		Россия	26 Да	
1 Мунской	130000	3 Неоконченное высшее	Профессионал	нет		3 0-1 KM	_	Espona	51 Да	
e a ar Krase	нты Табли	цаЗ Outlers Заполнение из п	ринера Прогнозир	Cf. suman				14		+

Рис. 4.3. Винятки виділяються кольором, що дозволяє провести сортування

На Рис. 4.2_видно, що інструмент дозволяє вказати поріг відхилення від норми (Exception threshold), вимірюваний у відсотках (оцінка ймовірності того, що виділене значення відноситься до винятків). Зменшення порога призведе до того, що більше записів буде розглядатися як виключення,

збільшення - навпаки. При значенні за замовчуванням в 75% нашому наборі даних виявлено 34 виключення. Звіт показує, в яких стовпчиках скільки винятків було виявлено.

Перейдемо на лист Excel з даними. Розглянуті як викиди значення виділяються в таблиці кольором: вся рядок-коричневим, конкретне значення - жовтим. Щоб згрупувати потрібні рядки можна скористатися функціями Excel, що дозволяють провести сортування за кольором.

Також можна скористатися інструментами вкладки "Вид", щоб створити нове вікно і розташувати поруч з вікном зі звітом і даними (Рис. 4.4). Нехай у відібраному наборі записів ми виявили ошібку.Скажем відстань до роботи у якогось клієнта з США, що володіє двома машинами, не «0-1 км", а "5-10 км" (саме тому йому потрібно в родині 2 машини). Якщо ми змінимо значення, буде проведений автоматичний перерахунок. У разі, представленому на Рис. 4.4, нове значення вже не розглядається як викид.

N A A A	Microsoft E	ice)	Print Income	Sec. No.
Charmente Information Parameter Information Parameter Information Parameter Information Parameter Information Information Parameter Information Paramete	annan Pepagina Jenne Pro Jerrite & Chenadeprop Grea & Jerroses Manurel Desarre	Integrations for a second seco	Programme inductor international programme internation	~ 0
17 * (* Ji 5-10	AM			-
Codepled50111_PLData_Mining_&conversional	eas2	o 8	a redepted NOIL /P. Data Merring, deptering research	
6 H	 1) 4) 	K L	A B C D E F	
1			Highlight Exceptions Report for Ta6лица3	
з зеления категорий", "Выде.	ления исключений" и "А	нализа сценариев"	3 The outlier cells are highlighted in the original table.	
2				
4 Тип работы 💌 Домовладелен 💌	вол во ило = Расстонные до раб	otsa 🖬 Person 💌 Bospact 💌	A Transford Barrishald Interest of Farman Antonia Contractional Table	
5 Профессионал Нет	3 0-1 «M	Tapona 31	a cospoon pretrois (more or rewer exceptions)	
в Профессионал Нет	201***	CUIA 49,	Colora	
7 Квалифицированный Да	2 5-10 eM	CILLA 44	A Constitute to restance	
 Кеалифицированный Де 	2 2-5 xM	CUIA 45		
 Квалифицированный Да 	D 0-1 #M	Espona 42	A format	
10 Офисный работник Да	1 0-1 «M	Eapona 43	10 Detu	
11 Профессионал Нет	2 2-5 KM	Espona 60	11 Official and 1	
12 Профессионал Да	1 5-10 KM	Pocces 41		
13 Офионый работных. Нет	0 0-1 «M	Espona 36		
3.4. Ручной труд Да	0 1-2 em	Espona 50		
15 Управление Де	4 0-1 kM	Poccest 33	S 10 Research an and second se	
28 Каалифицированный Да	0.0-1 xM	Espona 43	A 16 Among to patients	
17 Офисиый работник Да	2 5-10 KM	Poccan 58	A TO PERMIT	
18 Ручной труд Де	1 0-1 KM	Espona 48		
13 Квалифицированный Нет	2 1-2 KM	Pocces 54	A 10 Hore 23	
20 Профессионал Нет	4 10+ KM	POCOAR 35	4 <u>80</u>	
23 Профессионал Да	4 0-1 xm	Expona 55	H 100	
22. Офионый работник Да	1 1-2 KM	Espona 35	B 20	
23 Квалифицированный Нет	10-1 «M	POCDAR 45	8	
24 Ручной труд Да	10-1 ×M	Espona 38	A 10	
25 Офионый работник Нет	2 1-2 mm	Poccas 59	A (2)	
28 Офисный работник Да	0 0-1 «M	Expona 47	t N	
27 Офисный работник Да	1 1-2 mm	Espona 35	A Distance of the second se	
28 Офисный работник Ди	2 5-10 KM	Potové 55	A A	
28 Офисный работник Да	0.0-1 kM	Espona 36	A	
35 Provide research Ra	A 104-year	Rotras 15	A A A B Document Taliance) Outliers (bronnesses at maneta (Documentarian 2)	
Termes	The state of the s	11 1 1111 11 2 113		

Рис. 4.4. Виправлення помилкового значення

Зверніть увагу, що не тільки змінилася розфарбування рядки таблиці, а й відбулися зміни в звіті, що показує тепер наявність 33 винятків.

Автоматичний перерахунок працює тільки в тому випадку, якщо сесія роботи з аналітичними службами SQLServer залишається відкритою. Якщо таблиця Excel була закрита і знову відкрита, то автоматичного перерахунку не буде (потрібно знову провести аналіз).

Також в описах зазначається, що інструмент реагує тільки на зміни даних в діапазоні осередків, використовувався при навчанні. Якщо після початку роботи інструменту в кінець таблиці додати нові рядки, вони оцінюватися не будуть.

Як вже зазначалося вище, якщо потрібно розглядати тільки найбільш сильні викиди, можна збільшити значення порога відхилення і інструмент змінить оцінки відповідно до заданого значенням (Рис. 4.5).

Повторний запуск інструменту видалить результати попереднього аналізу. З огляду на, що проводяться інструментом зміни досить складні (розфарбування рядків таблиці і т.д.), якщо потрібно видалити результати роботи, рекомендується запустити повторний аналіз, погодитися з видаленням результатів і потім у вікні, аналогічному представленому на Рис. 4.1, Натиснути кнопку Close (відмовитися від аналізу даних).

X	<mark>, , , , , , , , , , , , , , , , , , , </mark>	deplex050	311_PI_Da	ta_Minir	пд_формат	ированая -	Micros		x
0	аил Главна Вставк;	Размет	Формул	Данны	Реценз	Бид Data I	И Рабоч		
		23	Q	唱					
06	ычный Разметка 🚃	Показать	Масштаб	Окно	Макросы				
Pere	страницы				Макросы				
	B5 •	. (m	f _x		THUR S CAL	1			
1		A				В	С	D	E
1			Highlig	tht Ex	ception	ns Repo	rt for	Таблиц	a3
2	The outlier cells are	highlight	ted in th	e origin	al table.				
3	ine outlier censure			e engin					
4	Exception threshold	(more o	rtewere	exception	ons)	83	<u> </u>		
6	Column				7 Out	tliers 🔽			
7	Семейное положен	ие				0			
8	Пол					0			
9	Доход					0			
10	Дети					5			
11	Образование					1			
12	Тип работы					1			
13	Домовладелец					0			
14	Кол-во авто					2			
15	Расстояние до рабо	ты				0			
10	Регион					0			
12	Приобрел велосили				_	0			
19	Итог					14			
20						- 4			
14	Криенты Т	36,000,03	Outliers	2200	лнение из	примера	1 4 L m		• []
For	080	аолицаз	Judicis	Joint			0% (-)	0	-

Рис. 4.5. Збільшення порога відхилення зменшує число винятків

Завдання. Проведіть аналіз винятків для набору даних "Клієнти" і значення порога в 90%. Запропонуйте інтерпретацію отриманих результатів.

Завдання 2. Проведіть аналіз винятків для набору даних "Прогнозування" (продажі по місяцях в різних регіонах). Запропонуйте інтерпретацію отриманих результатів.

Аналіз сценаріїв

Інструмент Scenario Analysis дозволяє моделювати вплив, який чиниться зміною одного з параметрів (значень одного стовпчика) на інший, пов'язаний з першим. В основі роботи інструменту лежить використання алгоритму Microsoft Logistic Regression. Для формування часової моделі потрібно навчальна вибірка, що містить не менше 50 записів [3].

Інструмент Scenario Analysis включає дві складові частини - "Аналіз сценарію пошуку рішень" (GoalSeek) і "Аналіз можливих варіантів" ("What-If").

"Аналіз сценарію пошуку рішень" (GoalSeek)

Використання інструменту GoalSeek дозволяє оцінити, чи зможемо ми досягти бажаного значення в цільовому стовпці, змінюючи значення обраного параметра. Інструмент дозволяє провести аналіз як для запису, так і для всієї таблиці.

Використовуючи цей інструмент треба бути готовим, що не для всіх варіантів запиту може бути отримана відповідь. Це може бути пов'язано з тим, що у вихідних даних немає цікавлять нас поєднань. Також можуть бути проблеми через типів даних.

Крім того, не можна забувати, що запит потрібно формувати з урахуванням знань про предметну область. Наприклад, можна запросити систему, якщо людина хоче збільшити річний дохід на 20 відсотків, чи треба йому купувати велосипед. І навіть отримати якусь відповідь. Але зрозуміло, що в такій постановці саме питання є безглуздим.

Нехай ми хочемо дізнатися, як буде впливати освіту на рівень достатку людини. Спочатку проведемо аналіз для запису. Наприклад, нас цікавить клієнт з ідентифікатором 12496 (перший запис в наборі даних). Откройте набір даних "Клієнти" і на вкладці Analysis виберіть ScenarioAnalysis-> GoalSeek (Рис. 4.6).



Рис. 4.6. Вибір інструмента GoalSeek

У вікні параметрів (Рис. 4.7) Вкажемо цільової стовпець - "Дохід", а також бажане значення - 125% від поточного. У цьому випадку інструмент вважає успішним результат, що не менше заданого (в нашому прикладі $40000 \times 1,25 = 50000$ і більше). Якщо задається значення менше 100%, то успішним вважається результат, що не більше заданого. Також можна вказати точне значення і діапазон (вибравши "Inrange"). Для значень, які не є числовими, варіанти "Percentage" і "Inrange" будуть неактивні. Для досягнення шуканого значення будемо міняти стовпець "Освіта".

1.1	10100 Store 100			Cuta tituinia	Part nicks minutes	Analyza Kon	- The second		
	Scenario Ana	lysis: Goal Seak	SQL Server 2008	- Children -	The second second	- sealer - we	- and a state		
Rey ani O	The Goal Seek tool provi the current row the Targe column) to reach that des your table.	des useful recommendations when you know the colume), and need to know how to change an ind value. The recommendations are based on	desired value for a column of other column the Change he patterns and rules detected in						
4	Geal to seek:			G	н	1	1	к	L
spie e	Tarpet	Донов	•		58 80. com				
MN				ния категори	и, выделя	ения исклю	чении и А	нализа сцен	ариев
	C Exectly:			-		Support of the local division of the local d		and planters of	and a second second
	· Percentage	125 O To of Desca		боты 🗖 до	worksteard .	104 BO ATTO - P	эсстовные до раб	onul - Petrion - I	ветраст Прие
190 1		100 million (1990)		рацированные да		00	F1 KM	Espona	42 Het
107 1	() in range:		3 1	Advanced Columns	Selection			tepora	45 Het
177	What to change:			This task analyzes th	e data in the table colu	anna to recommend o	olumns for use	Espona	41 00
100	Channe	Oferenant		in analysis, tou can columns below	overlide Pris recommer	station by manually c	noowing the	FREEDO	41 100
07	1.1.1			The second second				Espons	50 Her
974		Choose columns to be used for engine		1				Pocces	33 /04
164	Specify Row or Tab	4e		Choose column	to be used for an	alysis		Tenona	43 04
155	· On this row	C Entire table		V Ofpeconeve				Porces	SS Her
280 1				Доновладеле				Fedora	48 /la
173 2			Rat	Kon_eo asto			13	Pocces	54 fla
6974	Results		1.000	Patroni Patroni	pational			Poccus	36 Het
434 1				Вораст				Espona	55 Her
323				Covortpen serv	x(wheat		14	Espona	35 Да
542						~ 1	Cried 3	Россия	45 Дa
870 6						un l	Cartes	Eapona	38 ,Ca
316 (ый работных. Не	т	21	-2 KM	Россия	59 Да
610 1				ый работных Да	2	0.0	H1 KM	Espona	47 Her
103.4				A asken in		4.4	Auger	Canada	36. 19.0

Рис. 4.7. Налаштування параметрів для GoalSeek

Перейшовши за посиланням "Choose columns to be used for analysis", відзначимо, що при аналізі в розгляд не беремо стовпці "ID", "Діти", "Відстань

до роботи", "Придбав велосипед" Після закриття вікна "Advanced Columns Selection" варто ще раз перевірити настройки в секції "Goaltoseek" - іноді при переході між вікнами перемикач між "Exactly", "Percentage" і "Inrange" сбрасиваетсяв значення за замовчуванням ("Exactly")

Scenario An	alysis: Goal Seek	SQL Server2					
e Goal Seek tool pro e current row (the Ta lumn) to reach that d ur table.	vides useful recommendations when you rget column), and need to know how to o esired value. The recommendations are	u know the desired value for a column of change another column (the Change based on the patterns and rules detected					
Goal to seek:							
Target:	Доход	÷					
Exactly:							
Percentage:	125 🔷 % оf Доход						
🕑 In range:							
What to change:							
Change:	Образование	•					
	Choose columns to be used for a	malysis					
Specify Row or Ta	able						
On this row	C Entire table						
		Run					
Results		41 11					
Goal Seeking	ng for Доход has found a solution.						
Образование' = Нес	оконченное высшее						
connucince. very i							

Рис. 4.8. Результат аналізу для одного рядка - рішення знайдено

Результат аналізу, виконаного після натискання кнопки Run, представлений на Рис. 4.8. Для вибраного рядка знайдений шаблон, що рекомендує для параметра "Освіта" значення "Незакінчена вища". При цьому

рівень достовірності - Confidence (іноді верхня частина напису затирається, як на малюнку), оцінюється як дуже низький ("Very low").

Якщо перейти на наступний рядок і знову натиснути Run, отримаємо результат для нових даних (Рис. 4.9). В цьому випадку, відповідного рішення не було знайдено, і був запропонований найбільш близький варіант.

Scenario An	alysis. Goal Seek	SQL Server 20
he Goal Seek tool pro e current row (the Tar plumn) to reach that d pur table.	vides useful recommendations when you know the de get column), and need to know how to change anoth esired value. The recommendations are based on the	esired value for a column of her column (the Change patterns and rules detected i
Goal to seek:		
Target:	Доход	÷
Exactly:		*
<u>P</u> ercentage:	125 🚖 % оf Доход	
🕑 In r <u>a</u> nge:		
What to change:		
Change:	Образование	•
	Choose columns to be used for analysis	
Specify Row or Ta	able	
On this row	Entire table	
		Run
Results		20 90
🔥 Goal Seekir	ig for Доход has not found a solution.	
Closest match, 35955 среднее. соппоснос, very н	(119,85%), obtained by changing "Образование" to:	Неоконченное
A CONTRACTOR OF A CONTRACTOR O		

Рис. 4.9. Результат аналізу для одного рядка - рішення не знайдено

		60	deplex050311_PLD	lata_Manag Microsoft E	stel		Fational Crait Program	(C) (C) - 8
dala fassian Sc	naesa Patwerna	праницая Форм	пули Дантия	Рецензирование	Seul Data Mining	Рабочая группа	Analyze Kowctpykrop	a 🕢 e 🚽 i
Caleri	- 11 -	A' x' = =	* * =	OSupril	· N	I	Transis E.	A
Benamera J K K	9 * 🗉 * 🆄	· 🗛 · 🖉 🖉	# (* (*))	· · · · · · · · · · · · · · · · · · ·	условное форматиравания -	Pephanipseats Cher Kak talining * Aven	a Boopean - 2 - a person	Halma a Bagenens *
Nó -	(- f1		46-112-01			4.000M	The second secon	
6	н	10	1	ĸ	4	M	N	0
4 Ten pañonar -	Домовладелец	· Kon-so arro	Paccetomer p	о работы - Регнон -	Возраст • Приобр	en senocement . G	oal: Доход charges by 125.00 % 💌	Recommended Oбразование
5 Квалифицированны	∉ Да	14-14 Oct 14-15	0 0-1 KM	Espona	42 Het	10-10-10-10-10 (1)	0	Неоконченное высшее
6 Офисный работник	Да		10-1 xM	Espona	43 Het		0	Неоконченное среднее
7 Профессионал	Het		2 2-5 KM	Espona	60 Her		0	Неоконченное среднее
8 Профессионал	Да		1 5-10 emt	Poccies	41 Де		0	Несконченное высшее
9 Офисный работник	Her		0.0-1 xxx	Eapona	36 Да		0	Несконченное высшее
10 Ручной труд	Да		0-1-2 mAR	Eepona	50 Her		0	Бакалаер
11 Управление	Да		4 0-1 RM	Россия	33 Да		0	Неоконченное высшее
12 Квалифицированны	й Да		0 0-1 KM	Eepona	43 Да		0	Неоконченное высшее
13 Офисный работник.	да		2 5-10 KM	Россия	58 Her		0	Несконченное высшее
14 Ручной труд	Да		10-1 ния	Espona	48 Да		0	Неоконченное среднее
15 Квалифицированны	# Her		2 1-2 MM	Россия	54 Да		0	Бакалаер
16 Профессионал	Her		4 10+ KM	Россия	36 Her		0	Несконченное высшее
17 Професснонал	Да		4-0-1 xM	Espona	55 Her		0	Неоконченное среднее
18 Офисный работник	(Da		1 1-2 KM	Espona	35 Да		0	Неоконченное среднее
19 Квалифицированны	# Het		1.0-1 км	Poccies	45 Да		0	Неоконченное среднее
20 Ручной труд	Дa		1 0-1 xAN	Espona	35 Да		Ö	Несконченное высшее
······································	New BODAHENNE HO ISHO	нра Прогнори	C 104000	BARRIE	80. D+		fk	
otoes							100	

Рис. 4.10. Аналіз для всієї таблиці

А якщо в секції "Specify Rowor Table" встановити перемикач в "Entire table", то сценарії будуть пораховані для всіх рядків (Рис. 4.10). Результати будуть вказані в двох стовпчиках, доданих в вихідну таблицю. Для тих рядків, які відзначені хрестиком в червоному колі, відповідного бажаного сценарієм шаблону знайдено не було.

Завдання. Проведіть аналіз для окремого рядка і таблиці, аналогічний описаному вище. Прокоментуйте результати.

Примітка. Запуск процедури аналізу для ряду інших комбінацій стовпців (наприклад - цільової стовпець "покупка велосипеда" = "так", незалежна змінна - "відстань до роботи") призводить до помилки "Query (1, 50) Синтаксичний аналізатор Невірний синтаксис" value ". ", мабуть пов'язаної з некоректною обробкою деяких типів даних.

"Аналіз можливих варіантів" ("What-If")

Інструмент What-If дозволяє вирішити зворотну по відношенню до GoalSeek завдання: оцінити значення цільової змінної при певному зміні заданого параметра.

Наприклад, можна оцінити, як змінився б рівень доходу людини, якби підвищився його рівень освіти. Перейдемо на запис з ідентифікатором 12697 і запустимо інструмент: Scenario Analysis-> What-If. Зазначимо параметри сценарію: освіта змінюється на "Вища" і цільової стовпець "Дохід". Отриманий для рядка результат показує, що при зміні рівня освіти дохід може дещо зрости (початкове значення 90000, середнє значення для нового шаблону 104448). Але ступінь впевненості в прогнозі не надто висока.

Аналогічно до попереднього інструменту, подібний аналіз сценарію можна зробити і для всієї таблиці цілком. В цьому випадку до вихідної таблиці додаються два стовпці - один показує нове значення цільового параметра, другий - оцінку достовірності (Рис. 4.12). Достовірність оцінюється числом від 0 до 100: 100 - максимальна достовірність (абсолютна впевненість в прогнозі), 0 - мінімальна.

Завдання. Проведіть аналіз даних, аналогічний описаному вище.

Для того щоб видалити результати роботи з таблицею інструментів What-If i Scenario Analysis, досить видалити додані стовпці. При роботі з окремими рядками, ніяких додаткових дій не потрібно.

Tabuar Erabas Tabuar Erabas Salyse Key Deted, Fill Prom For forences Categories Stample	Panartia reparato	Degraphie Degraphie	Derenet Prodestopolaseet Sealogy Children Die Bester Analysis Competition	? Scener	io Analysis: Whet If makates the inpact that changes in one colorer re-	SQL Server 2008	6 0 0 # 1
A16 .	6 12697			a parameter	THE CAPET IN IN WITH PERFORME	1	1
INAME IN CONTRACTOR		0		Scenario			1 10
	-			Charce	Oterman		-
D Complete pagaret	weet 🖬 Data 🔤	Downes - Des	Te Constanting				astaler 💌 New Open av atcente a
12496 Женатый, замужий	Anacoust .	40000	1 Eaxanaep		To salue: Bucune		42 Het
24107 Menatore, sanyment	Myapipä	30000	3 Hegebievening autoure		Colorenzation (OII - 64)		43 Her
14177 Женатый, замужий	Mysessall	80000	5 Неоконченное высшее	20			50 Her
24381 Opencepil(an)	Myspecie	70000	0 Gexanaep	What happe	ns to		43.04
25597 Ogenovnik(an)	Myeckoli	30000	0 Бакалаво		(mark)		36 (24
3 13507 Женатый, замужняя	Женский	10000	2 Heastweener Bucuee	range.	ABROA		50 Her
1 27974 Одиноний(ан)	Мужской	160000	2 Cpeakee		Choose only may be be used for analysis		33 <i>Ca</i>
2 19364 Женатый, замужий	Mywpepä	40000	1 Бакалаво	10	Accession of the second second second		43 (54
22155 Menature, sawyweek	MyxecxcA	20000	2 Неоконченное среднее	de Specify Row	or Table		58 Her
19280 Женатый, замужняя	Mysecsol	20000	2 Неоконченное высшее	Co this row	C Entre sole		45 <u>Д</u> а
5 22173 Henatsik, sawywow	Silenconia	30000	3 Средное	14			54 Da
12697 Одинский(ая)	Менский	90000	0 Eakanaep	100		- Page -	35 Her
11434 Женатой, замужно	Мужрхой	170000	5 Неоконченное высшее	20			55 Her
8 25323 Женатый, замужние	Мужрюй	40000	2. Неоконченное выпшее	Del Parales			35 Да
23542 Одиноний(ая)	Mysocial	60000	1 Неоконченное высшее	the S What	If Analyzia for Exception transfer solution.		45 <u>Д</u> а
20870 Одинокий(ая)	Женский	10000	2 Среднее		Management and a second second		38 <i>Д</i> а
23316 Одиноний(ая)	MyxebioiP	30000	3 Неоконченное высшее	De Zinoz + 104	646 <u>8</u>		59 Дa
2 12610 Женатый, замужию	Менский	30000	1 Eaxanasep	0. Cutter	Mary Low		47 Her
3 27183 Оденокий(ая)	Мужской	40000	2 Несконченное висшее	0.			35 Да
1 25940 Ogenowsk(an)	Мужриой	20000	2 Неоконченное среднее	04			55 <u>A</u> a
5 25598 Meleature, santymente	Mentsal	40000	0 Bucupee	0			36 jūja

Рис. 4.11. Робота інструменту What-Іfдля окремого рядка

1 A 494 C	1.1	80	deplex090101_PLData_Mar	ing - Mcrosoft Excel	100 million (1990)		Patients Californ	in the second second	_	-	Col (C)
main from	Hart Bittabl	в Расметка праницая фере	una dannat Pru	exceptioner Eng	Deta Mining	Parlionan rayon	a Analyze Kow	труктор			
n 4.	Califies	$ \cdot _{W} \cdot _{X',X'} = =$	Britepere	N 199074	Chard		<u>11</u>	18 19	3- 3- 3	x - 3	e an
. 1	X X Y	· <u>.</u> . <u>.</u> . <u>.</u> . <u>.</u> .	a to to Moore	worth a substitute a file		001 75 47	Construction . 13	a tablang * avers	Bankers	2. s que	170 * Bulletors *
05		E IN TANKS LIKENT								- Copie	
		A PERMIT		67		27	1.20	1	22		
a c	D	E 7.	0.	16.		1.			M	N	0
Dot 1	Concerning and	a commune	Destandaria -	Doutonal and - D	and a second at 12	And the second second	Internet a Deserve a	Bernard a Dree	Costs an appropriate in the	- 17	and the second second
Westman	40000	1 Envenance	Kas naha umpassion	/Ta	0.0	1.244	Teorina	67 Her		56575	
Montana	10000	3 Hereolanaturative Institute	Otwowii nafotnez	/a	1.0	tax	Famona	43 Her		32585	
Montacia	80000	5 Mechanisments particulat	Поофестночна	Hot	22	-S also	Espona	60 Her		68975	-
Myerxold	20000	O Eavadage	Поофессионал	Z 4	15	-10 am	Porcess	41.04		67304	
Myaraoa	30000	0 Eavenage	Офирный работник	Her	0.0	-1 whi	Escona	35 (24		38673	
0 Mitheaching	10000	2 Неоконченное высшее	Ручной труд	fia.	01	-2 x M	Espona	50 Het		17296	-
I Мужской	150000	2 Cpequee	Управление	Da.	40	-I RM	Pocces	33 (24		125621	
2 Мужской	40000	1 Eavenage	Кеалифицированный	dia .	0.0	-1 #M	Espona	43 file		63024	
Mysicioa	20000	2 Несконченное среднее	Офисный работник	(La	2.5	-10 KM	Poccas	58 Her		22301	
4 Мужской	20000	2 Несконненное высыее	Ручной труд	Ди	10	-1 mm	Espona	48,004		22649	
5 женский	30000	3 Средное	Квалифицированный	Het	2.1	-2 mM	Россия	54 Д а		55595	
6 женский	90000	0 EakaAsep	Профессионал	Het	41	0+ a.m.	Pocces	36 Her	5	104445	
7 Мужской	170000	5 Неоконченное высшее	Профессионал	A4	4.0	-1 KM	Tapona	55 Het		106317	
8 Мужской	40000	2 Неоконченное высшее	Офисный работник	Да	11	-2 RM	Espona	35 JQA	2	38592	
9 Мунской	60000	1 Неоконченное высшее	Квалінфицированный	Her	1.0	-1 KM	Pocces	45 "Ilja		60032	
4 8 81 Kras	BATH / BOOM	Nonverse formers formers	Contraction of the second	-			10210141	D+			1
			Contraction of the second						140	11 100% (m)	0-0-0

Рис. 4.12. Прогноз What-Іfдля всієї таблиці