

What is Data Mining?



Data Mining

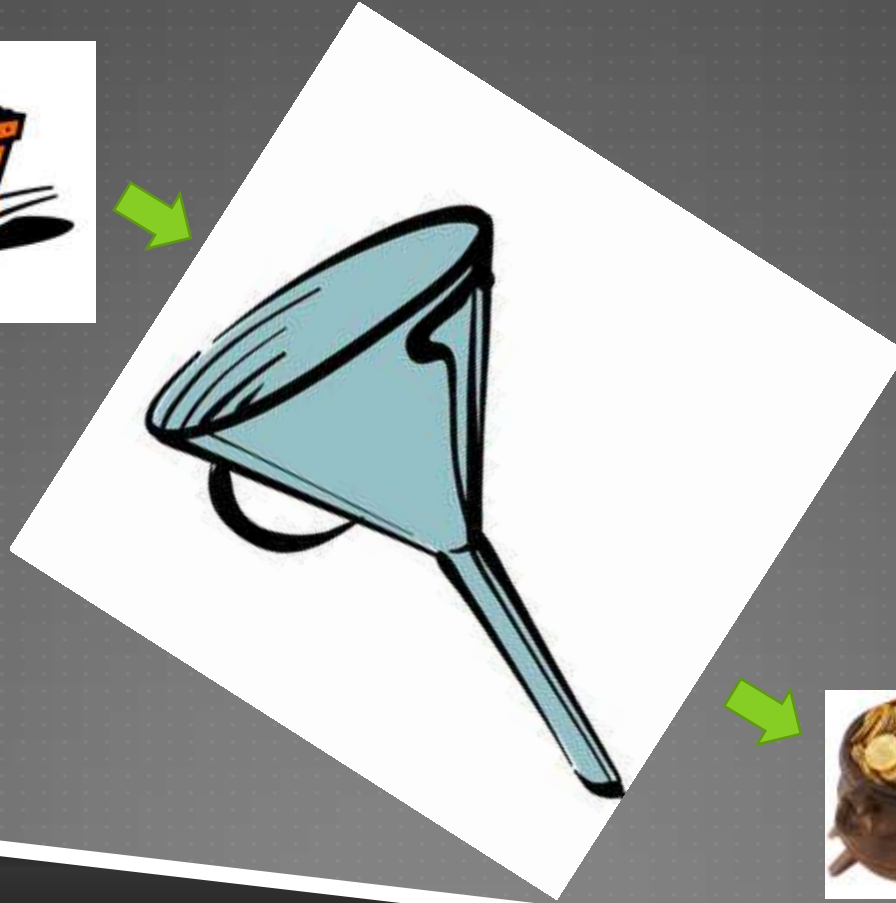
“The process of discovering meaningful new relationships, patterns and trends by sifting through data using pattern recognition technologies as well as statistical and mathematical techniques.”

- The Gartner Group

Data Mining



Data in



Info out

FRAUD DETECTION METHODS

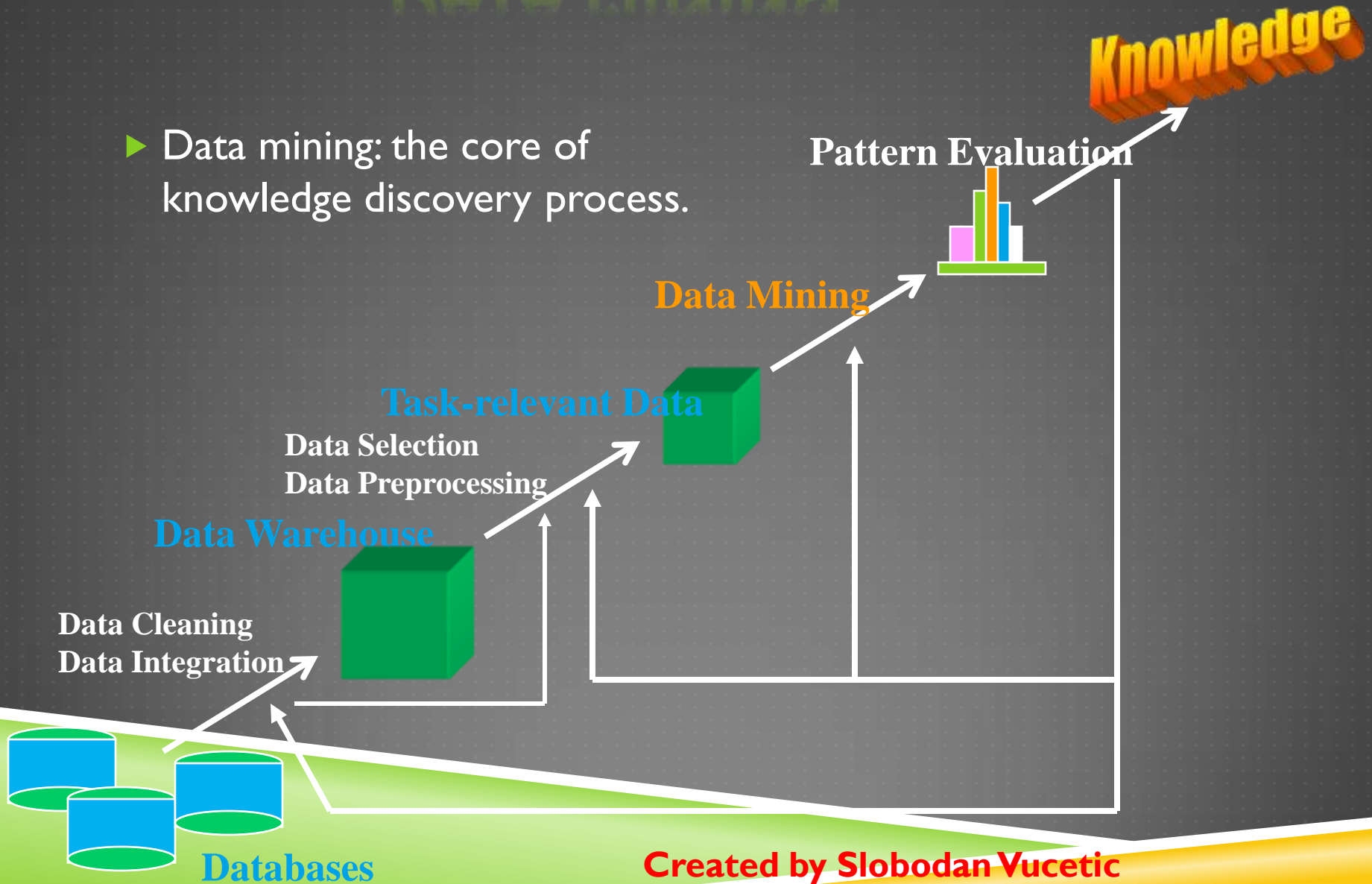


DATA MINING

- Fraud hotline
- Thresholds/alerts set up in ACL or Excel like products
- Forensic Accounting
- Data mining

DATA MINING

- ▶ Data mining: the core of knowledge discovery process.

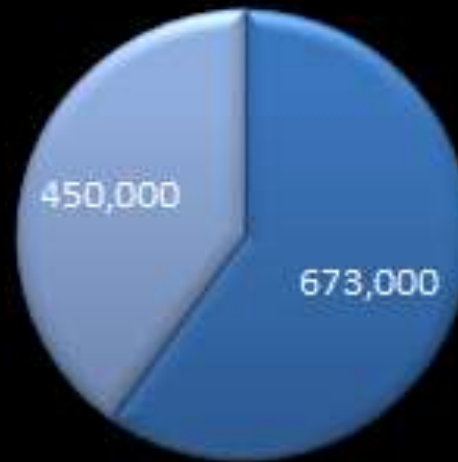


Created by Slobodan Vucetic

December 2012 Google Search Stats for "WIC"

Number of "WIC" Google Searches

(For the Month of Dec 2012)



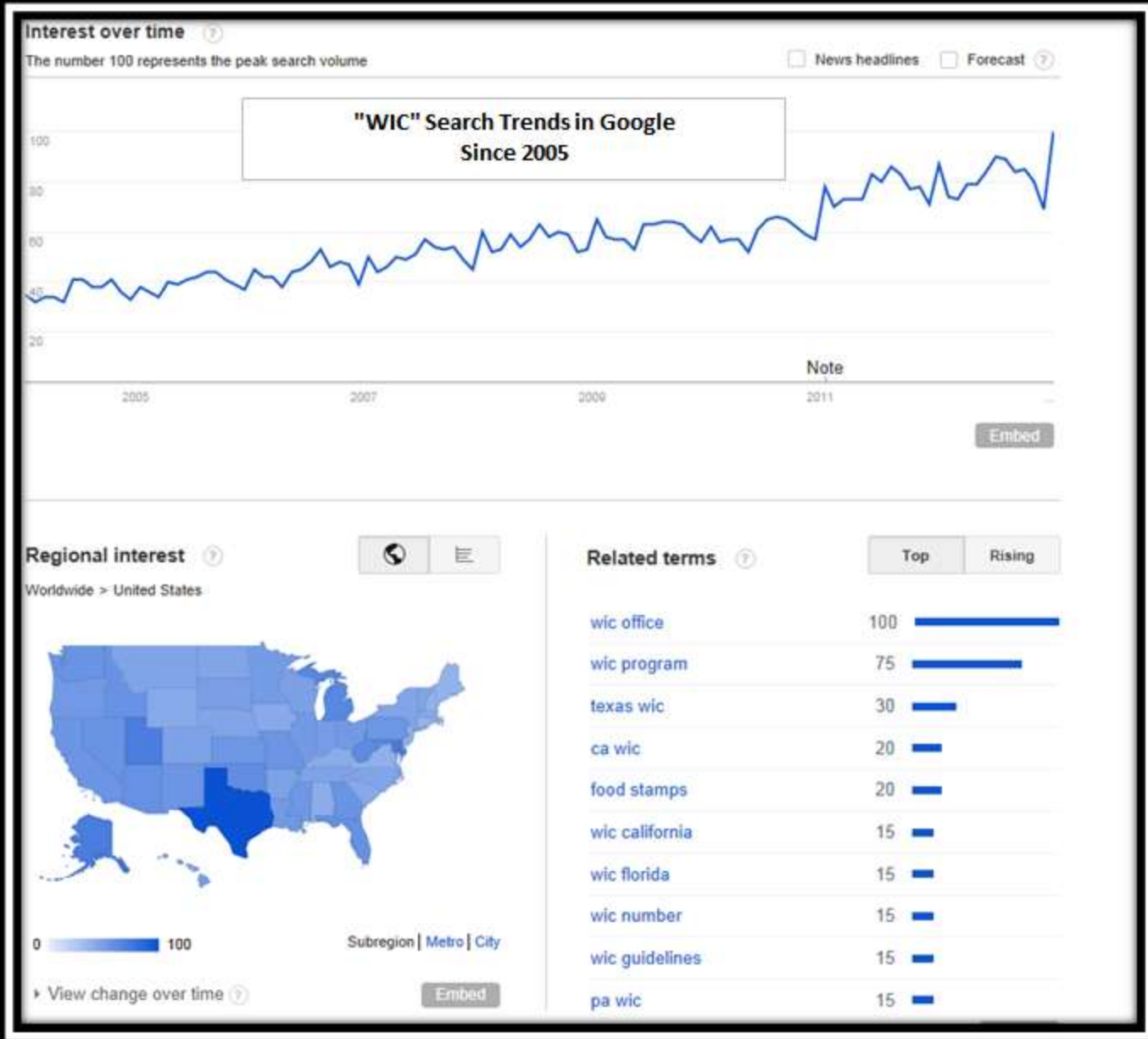
■ World Wide Searches

■ United States Searches

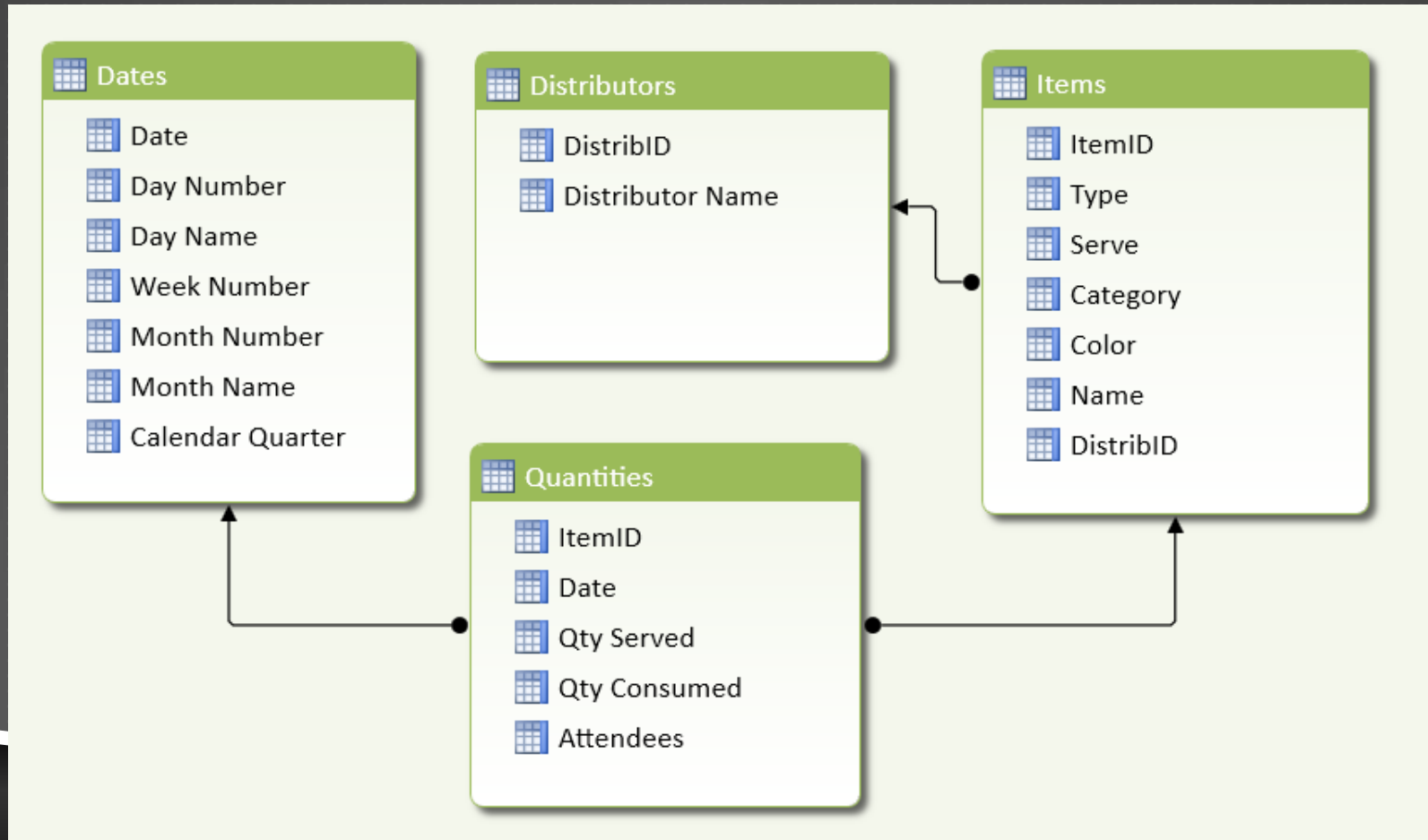
Is this Data Mining?

Is This

Data Mining?

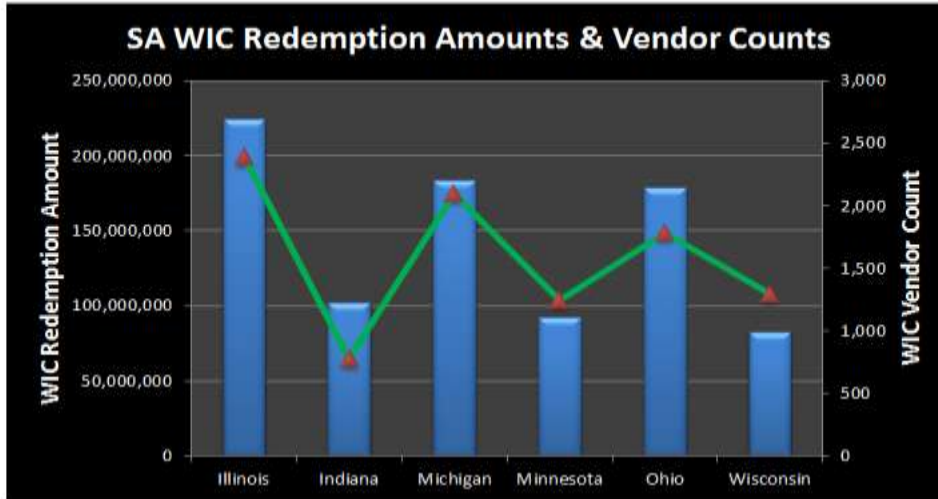


IS THIS DATA MINING (CONT.)



Excel Power Pivot addin & FNS Summer Food Program Dashboard

IS THIS DATA MINING?



Fiscal Year

- 2009
- 2010
- 2011
- 2012

Region

- Mid Atlantic
- Midwest
- Mountain Plains
- Northeast
- Southeast
- Southwest
- Western

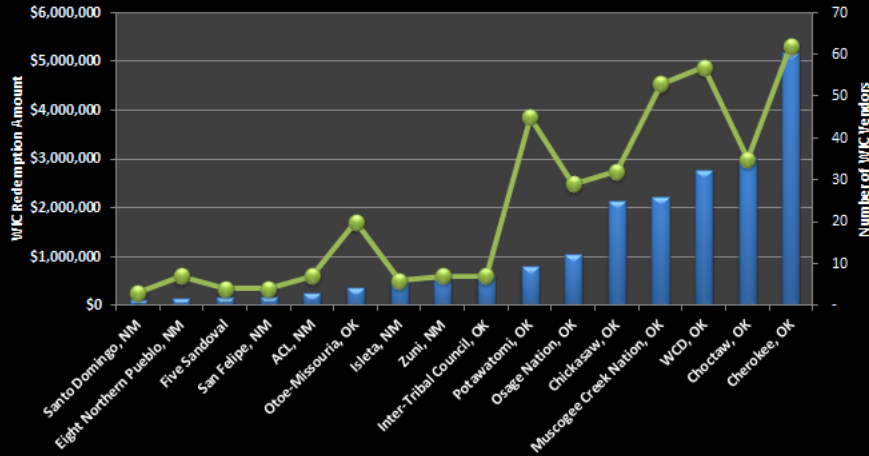
State Agency

Illinois	Indiana	Michigan	Minnesota	Ohio	Wisconsin	ACL, NM	Alabama	Alaska	American ...
Arizona	Arkansas	California	Cherokee,...	Cherokee,...	Cheyenne...	Chickasa...	Choctaw, ...	Choctaw, ...	Colorado
Common...	Connecticut	Delaware	District of...	Eastern S...	Eight Nort...	Five Sand...	Florida	Georgia	Guam
Hawaii	Idaho	Indian To...	Inter-Trib...	Iowa	Isleta, NM	ITCA, AZ	ITCN, NV	Kansas	Kentucky
Louisiana	Maine	Maryland	Massachu...	Mississippi	Missouri	Montana	Muscoge...	Navajo Na...	Nebraska
Nevada	New Ham...	New Jersey	New Mexi...	New York	North Car...	North Da...	Northern ...	Oklahoma	Omaha, NE
Oregon	Osage Na...	Otoe-Mis...	Pennsylva...	Pleasant ...	Potawato...	Puerto Rico	Rhode Isl...	Rosebud, ...	San Felipe...
Santee Sl...	Santo Do...	Seneca N...	South Car...	South Da...	Standing ...	Tennessee	Texas	Three Affil...	Utah
Ute Mou...	Vermont	Virgin Isla...	Virginia	Washingt...	WCD, OK	West Virg...	Winnebago	Wyoming	Zuni, NM

Slicers & TIP Visual Front End PPT

IS THIS DATA MINING?

SWRO WIC Vendor Redemption & Count Select 2009 to 2012



State Agency

ACL, NM	Arkansas	Cherokee, OK
Chickasaw, OK	Choctaw, OK	Eight Northern Pu...
Five Sandoval	Inter-Tribal Council...	Isleta, NM
Louisiana	Muscogee Creek N...	New Mexico
Oklahoma	Osage Nation, OK	Otoe-Missouria, OK
Potawatomi, OK	San Felipe, NM	Santo Domingo, NM
Texas	WCD, OK	Zuni, NM

Inventory Audits

0

1

2

(blank)

Admin Review Stat...

None

Overtuned

Pending

Upheld

Vendor Agreement ...

None

Terminated

Expired

Year

2009

2010

2011

2012

WIC Redemption Mon...

8

10

11

12

1

6

9

Indicator Name

A; C; H

B; H; X

E; X

F; G; L; V; X

F; H; X

F; X

G; H; X

G; X

H

Compliance Investi...

Completed

Initiated

None

On-going

No Sanction

N

W

(blank)

Routine Monitoring...

0

1

2

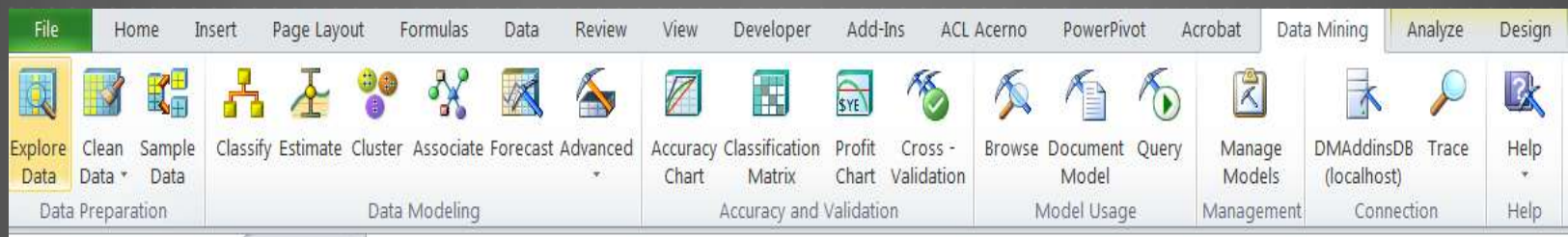
3

New Vendor

No

2013_TIP_WIC_44_Slicers in Excel

MICROSOFT SQL SERVER 2008 DATA MINING ADD-INS



SQL/Excel Data Mining Techniques

- ▶ Classification
- ▶ Clustering
- ▶ Regression
- ▶ Association Rules

ANALYZE KEY INFLUENCERS

Table Analysis Tools

	A	B	C	D	E	F	G	H
1	Sample data for Analyze Key Influencers, Detect Categories, Highlight Exceptions							
2								
3	ID	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner
4	12496	Married	Female	40000	1	Bachelors	Skilled Manual	Yes
5	24107	Married	Male	30000	3	Partial College	Clerical	Yes
6	14177	Married	Male	80000	5	Partial College	Professional	No
7	24381	Single	Male	70000	0	Bachelors	Professional	Yes
8	25597	Single	Male	30000	0	Bachelors	Clerical	No
9	13507	Married	Female	10000	2	Partial College	Manual	Yes
10	27974	Single	Male	160000	2	High School	Management	Yes
11	19364	Married	Male	40000	1	Bachelors	Skilled Manual	Yes

ANALYZE KEY INFLUENCES (CONTINUED)

Key Influencers Report for 'Purchased Bike'

Key Influencers and their impact over the values of 'Purchased Bike'

Filter by 'Column' or 'Favors' to see how various columns influence 'Purchased Bike'

Column	Value	Favors	Relative Impact
Cars	2	No	
Marital Status	Married	No	
Region	North America	No	
Cars	0	Yes	
Marital Status	Single	Yes	
Cars	1	Yes	
Region	Pacific	Yes	

Discrimination between factors leading to 'No' and 'Yes'

Filter by 'Column' to see how different values favor 'No' or 'Yes'

Column	Value	Favors No	Favors Yes
Cars	2		
Cars	0		
Marital Status	Married		
Marital Status	Single		
Cars	1		
Region	Pacific		
Region	North America		

Influencers: Marital Status?
Gender?
of Cars?
Region?

TIP DATA ALL SAS 2009 TO 2012

Key Influencers Report for 'Risk Designation'			
Key Influencers and their impact over the values of 'Risk Designation'			
Filter by 'Column' or 'Favors' to see how various columns influence 'Risk Designation'			
Column	Value	Favors	Relative Impact
Indicator Name	Missing	Non High-Risk	
Compliance Investigation Status	None	Non High-Risk	
No Sanction	Missing	Non High-Risk	
Sanction Reason	Missing	Non High-Risk	
Compliance Buys	Missing	Non High-Risk	
Sanction Type	Missing	Non High-Risk	
Inventory Audits	Missing	Non High-Risk	
Other Investigation Status	None	Non High-Risk	
Region	Mountain Plains	Non High-Risk	
Compliance Investigation Status	Completed	High-Risk	
No Sanction	N	High-Risk	
Compliance Buys	2	High-Risk	
Indicator Name	V	High-Risk	
State Agency	New York	High-Risk	
Sanction Reason	V	High-Risk	
Compliance Investigation Status	On-going	High-Risk	
No Sanction	W	High-Risk	
Compliance Buys	3	High-Risk	
Region	Northeast	High-Risk	
Indicator Name	W	High-Risk	
Indicator Name	F	High-Risk	

TIP Data Mining.xlsx Analyzing Key Influencers for Risk Designators

DETECT CATEGORIES & TIP VENDOR DATA

12 categories were detected

To rename a category, edit the 'Category Name' below.

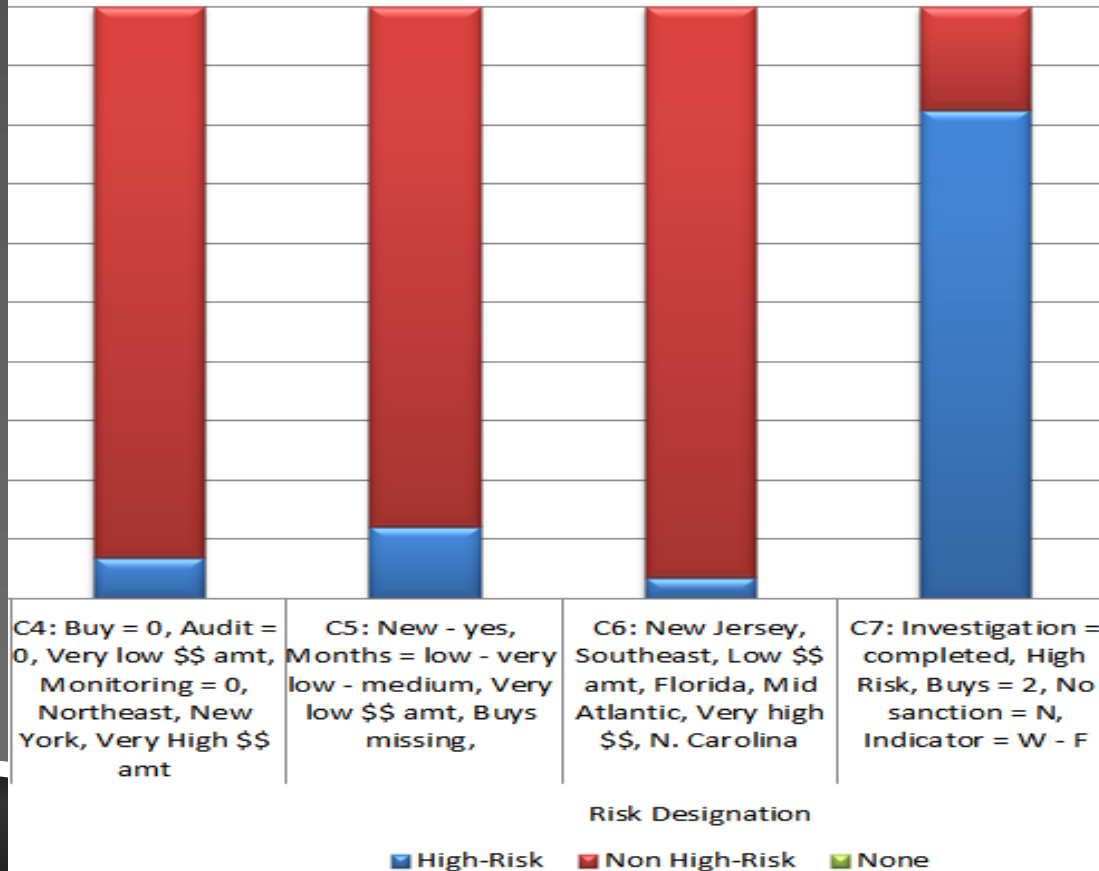
('Category Name' changes are visible in the 'Category' column of the source Excel table)

Category Name	Row Count
C1: Low \$\$ amt, Buy missing, Inventory missing, Midwest, Illinois	38,327
C2: Buy missing, Audit missing, Monitoring missing, Pennsylvania, Very high \$\$ amt.	20,137
C3: Buy missing, Southwest, Inventory missing, Texas, Mountain Plain Very high \$\$ amt	23,872
C4: Buy = 0, Audit = 0, Very low \$\$ amt, Monitoring = 0, Northeast, New York, Very High \$\$ amt	21,125
C5: New - yes, Months = low - very low - medium, Very low \$\$ amt, Buys missing,	16,219
C6: New Jersey, Southeast, Low \$\$ amt, Florida, Mid Atlantic, Very high \$\$, N. Carolina	16,251
C7: Investigation = completed, High Risk, Buys = 2, No sanction = N, Indicator = W - F	14,508
C8: Southeast, N Carolina, Very low \$\$ amt, Florida, Buys = M, Audits = M, Tennessee	11,755
C9: Audits = 0, Northeast, Buys = 0, New York, New Vendor = Y, Very Low \$\$ amt	9,394
C10: Months = 8 to 11, Sanction = V, High Risk, Southwest, Investigation = Completed	8,200
C12: Buys = 0, Audits = 0, New York, Northeast, Arizona, Low \$\$ amt, Monitoring = 0	8,344
C13: Southeast, Kentucky, Georgia, Indicator = M, Monitoring = 0, Audits = 0, Buys = 0	8,989
	197,121

TIP Data: Detect Categories

Category profiles (distributions of column values in categories)

Category and column. **Update Ctl, Alt & F5**



FUTURE DATA MINING NEEDS



THANK YOU!

Kelly Yee

Kelly.yee@fns.usda.gov

Tony Hardy

Tony.hardy@fns.usda.gov

Sunil Saraf

Sunil.saraf@fns.usda.gov

William (Bill) VanHouten

William.vanhouten@fns.usda.gov

