



IBM **Information Management** software

IBM WebSphere Information Analyzer Profiling and Analysis Reports



GlobalCo Sample Reports



IBM® WebSphere® Information Analyzer Profiling and Analysis Reports

GlobalCo Sample

The following reports were generated using WebSphere Information Analyzer for a fictional company called GlobalCo. The charts show a range of functionality for WebSphere Information Analyzer and the type of reports that are most commonly generated.

For more information about WebSphere Information Analyzer, call your IBM representative or visit us online at ibm.com/software/data/integration/information_analyzer

Table of Contents



Column Analysis—Data Types	4
Column Values—Most Frequent Values	5
Column Analysis—Frequency	12
Column Analysis—Most Frequent Formats	13
Column Analysis—Summary Statistics	14
Column Analysis—Structural Summary	15
Domain Analysis—Column Domain Values	16
Domain Analysis—Column Domain Values Detail	17
Column Analysis—Column Properties (by table)	18
Column Analysis—Column Data Types (by table)	19
Primary Key Analysis—Candidate Duplicate Exceptions	20
Table Analysis—Primary Key Chosen	21
Table Analysis—Primary Key Candidates	22

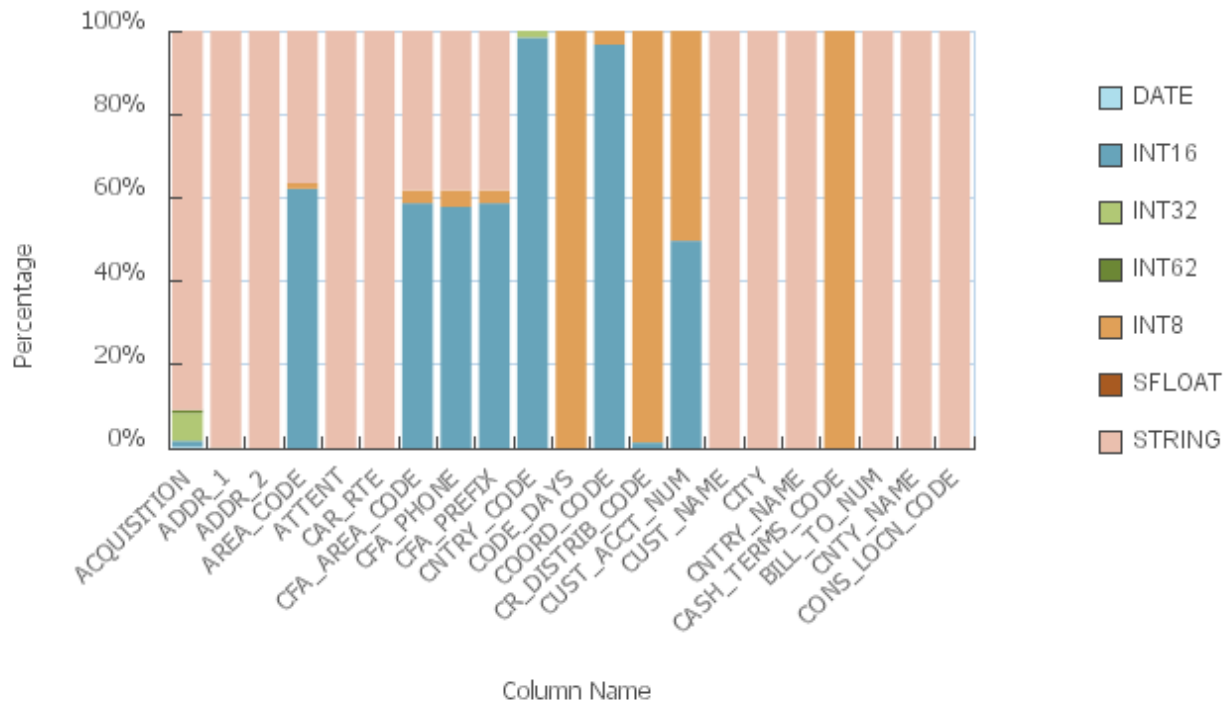
Column Analysis—Data Types

Column Properties	
Column All Data Types	
Product Module :	WebSphere Information Analyzer
Project :	GlobalCoIntegration
Report Name :	Column All Data Types
Date/ Time Executed :	03/01/07 12:46:02
Generated By :	admin admin
Customer Description :	GlobalCo ShipTo - Data Type Review
Host Name :	IBM-011956133F7
Data Store :	SQLServerE
Data Store Alias :	
Database Name :	dbo
Table Name :	GlobalCoShipTo
Table Alias :	



Column Summary

Candidate Data Type Distribution



Column Values—Most Frequent Values

Column Frequency

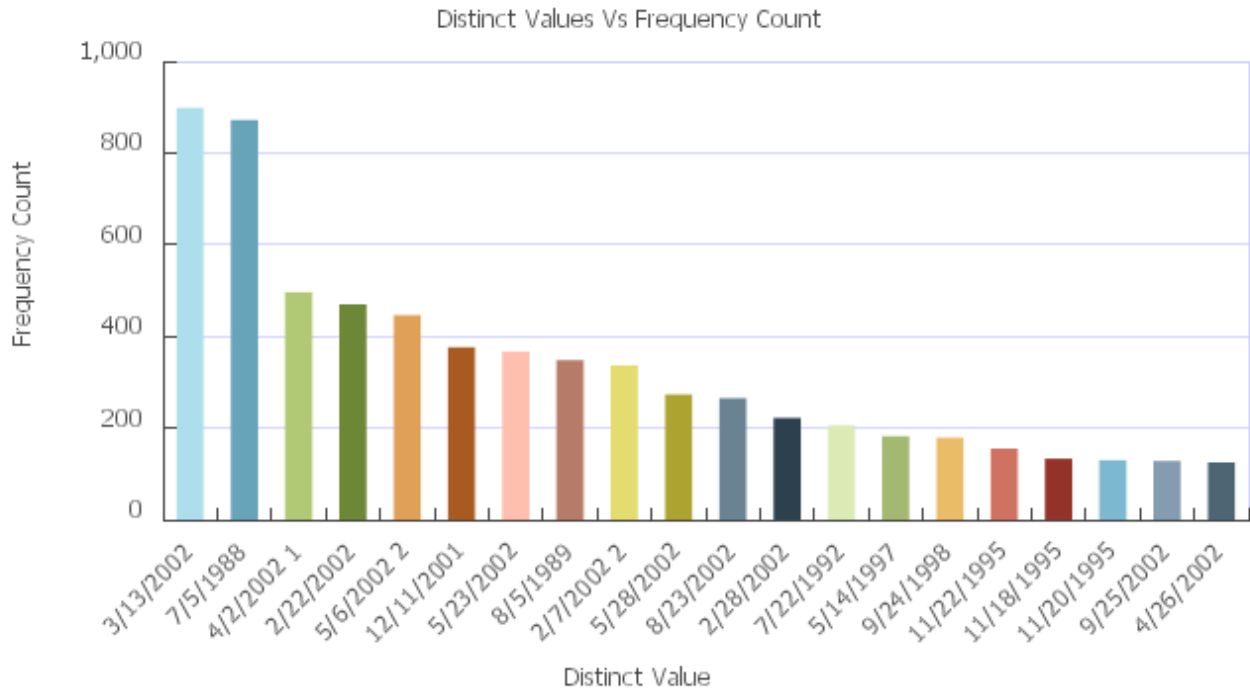
Most Frequent Values



Host Name :	IBM-011956133F7
Data Store :	SQLServerE
Data Store Alias :	
Database Name :	dbo
Table Name :	GlobalCoShipTo
Table Alias :	
Column Name :	SETUP_DATE

Column Level Summary

Cardinality Count :	6404
Null Count :	0
Actual Row Count :	37712
Total Rows Covered :	10364
%Rows Covered :	27.48





Column Values—Most Frequent Values (Cont.)

Frequency Distribution Data			
Distinct Value	Frequency Count	Frequency %	Cumulative %
3/13/2002	898	2.38	2.38
7/5/1988	872	2.31	4.69
4/2/2002 1	496	1.32	6.01
2/22/2002	470	1.25	7.25
5/6/2002 2	446	1.18	8.44
12/11/2001	376	1.00	9.43
5/23/2002	367	.97	10.41
8/5/1989	348	.92	11.33
2/7/2002 2	337	.89	12.22
5/28/2002	273	.72	12.95
8/23/2002	265	.70	13.65
2/28/2002	222	.59	14.24
7/22/1992	206	.55	14.79
5/14/1997	182	.48	15.27
9/24/1998	179	.47	15.74
11/22/1995	155	.41	16.15
11/18/1995	133	.35	16.51
11/20/1995	130	.34	16.85
9/25/2002	128	.34	17.19
4/26/2002	125	.33	17.52
11/19/1992	109	.29	17.81
11/30/1995	107	.28	18.10
8/13/1999	103	.27	18.37
3/1/1994 1	98	.26	18.63
11/21/1995	97	.26	18.89
2/23/1994	93	.25	19.13
12/21/1995	92	.24	19.38
9/19/1995	91	.24	19.62
9/13/1995	90	.24	19.86
8/11/1999	88	.23	20.09
10/17/1995	87	.23	20.32
10/8/1995	84	.22	20.54
8/22/1989	83	.22	20.76
4/10/2002	81	.21	20.98
9/25/1992	80	.21	21.19
10/27/1992	77	.20	21.39
7/2/2002 1	77	.20	21.60
10/19/1992	76	.20	21.80
3/2/1994 7	75	.20	22.00
2/25/1998	73	.19	22.19
1/24/1998	72	.19	22.38
11/28/1995	72	.19	22.57
7/13/1995	72	.19	22.76

Column Values—Most Frequent Values (Cont.)

Column Frequency

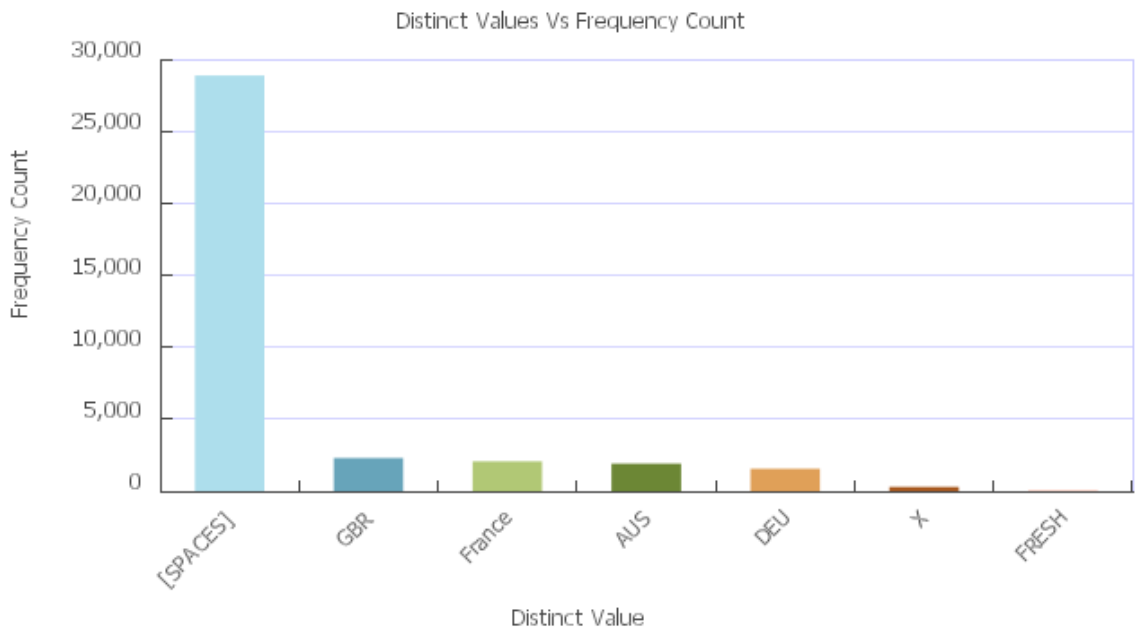
Most Frequent Values



Host Name :	IBM-011956133F7
Data Store :	SQLServerE
Data Store Alias :	
Database Name :	dbo
Table Name :	GlobalCoShipTo
Table Alias :	
Column Name :	CNTRY_NAME

Column Level Summary

Cardinality Count :	436
Null Count :	0
Actual Row Count :	37712
Total Rows Covered :	37059
%Rows Covered :	98.27



Frequency Distribution Data

Distinct Value	Frequency Count	Frequency %	Cumulative %
[SPACES]	28868	76.55	76.55
GBR	2306	6.11	82.66
France	2069	5.49	88.15
AUS	1916	5.08	93.23
DEU	1552	4.12	97.35
X	293	.78	98.12
FRESH	55	.15	98.27

Column Values—Most Frequent Values (Cont.)

Column Frequency

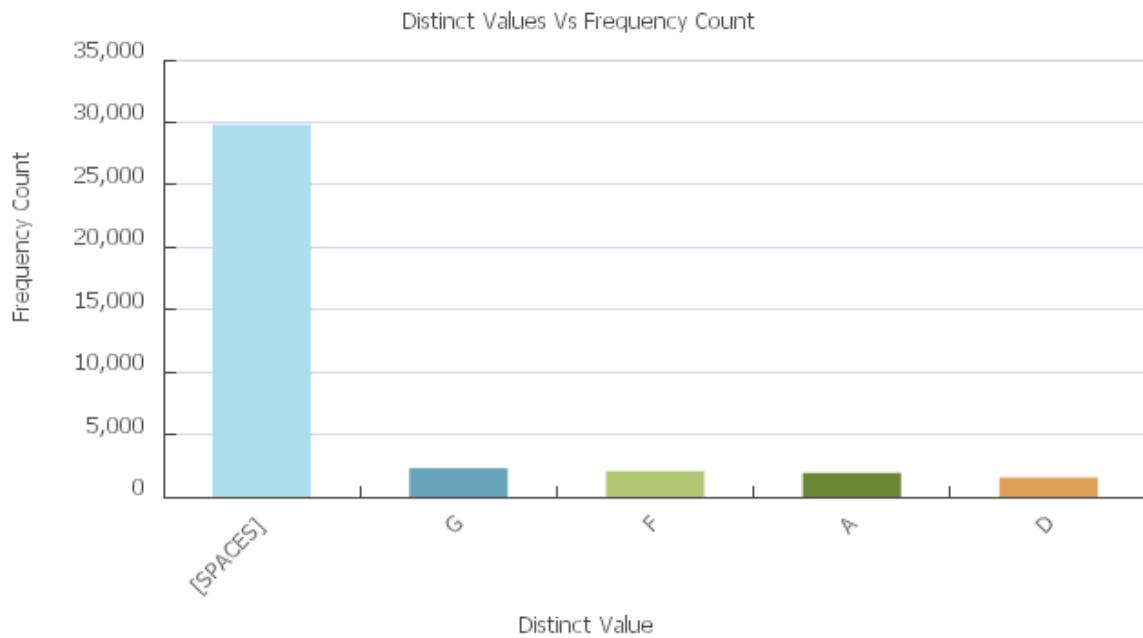
Most Frequent Values



Global Company

Host Name :	IBM-011956133F7
Data Store :	SQLServerE
Data Store Alias :	
Database Name :	dbo
Table Name :	GlobalCoShipTo
Table Alias :	
Column Name :	CNTY_NAME

Column Level Summary	
Cardinality Count :	10
Null Count :	0
Actual Row Count :	37712
Total Rows Covered :	37680
%Rows Covered :	99.92



Frequency Distribution Data			
Distinct Value	Frequency Count	Frequency %	Cumulative %
[SPACES]	29835	79.11	79.11
G	2306	6.11	85.23
F	2069	5.49	90.71
A	1918	5.09	95.80
D	1552	4.12	99.92

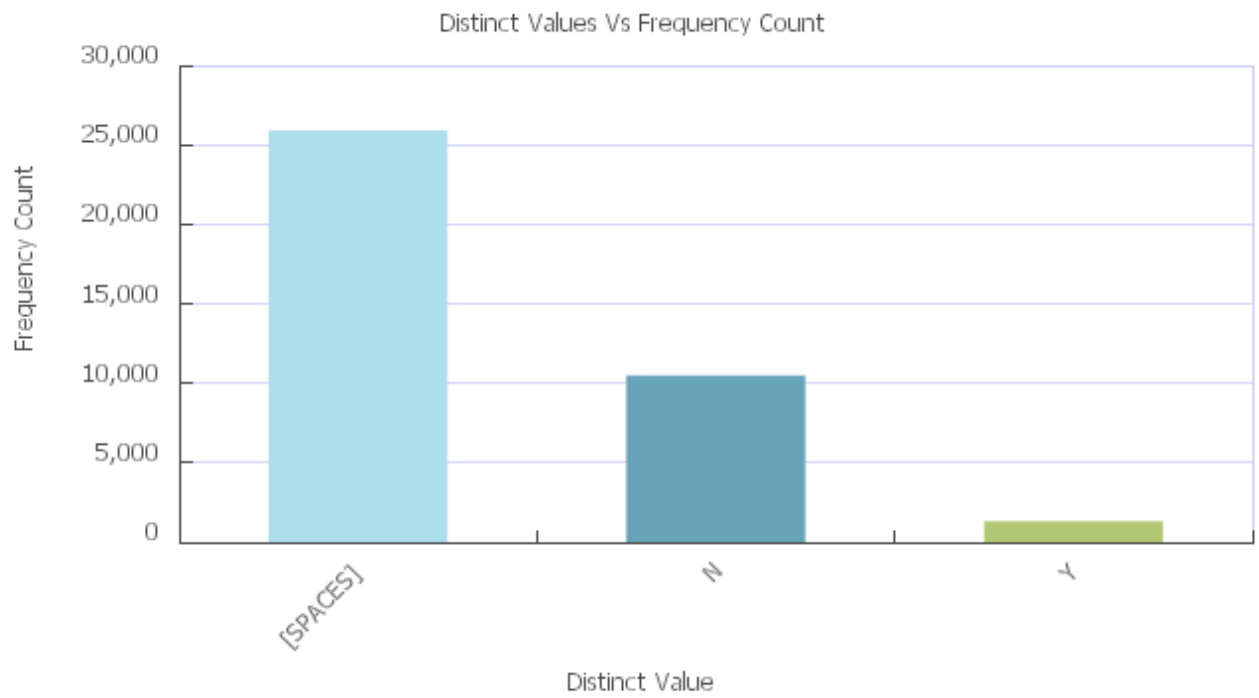
Column Values—Most Frequent Values (Cont.)

Column Frequency

Most Frequent Values

Host Name :	IBM-011956133F7
Data Store :	SQLServerE
Data Store Alias :	
Database Name :	dbo
Table Name :	GlobalCoShipTo
Table Alias :	
Column Name :	INTL_CUST_FLAG

Column Level Summary	
Cardinality Count :	3
Null Count :	0
Actual Row Count :	37712
Total Rows Covered :	37712
%Rows Covered :	100.00



Frequency Distribution Data			
Distinct Value	Frequency Count	Frequency %	Cumulative %
[SPACES]	25907	68.70	68.70
N	10480	27.79	96.49
Y	1325	3.51	100.00

Column Values—Most Frequent Values (Cont.)

Column Frequency

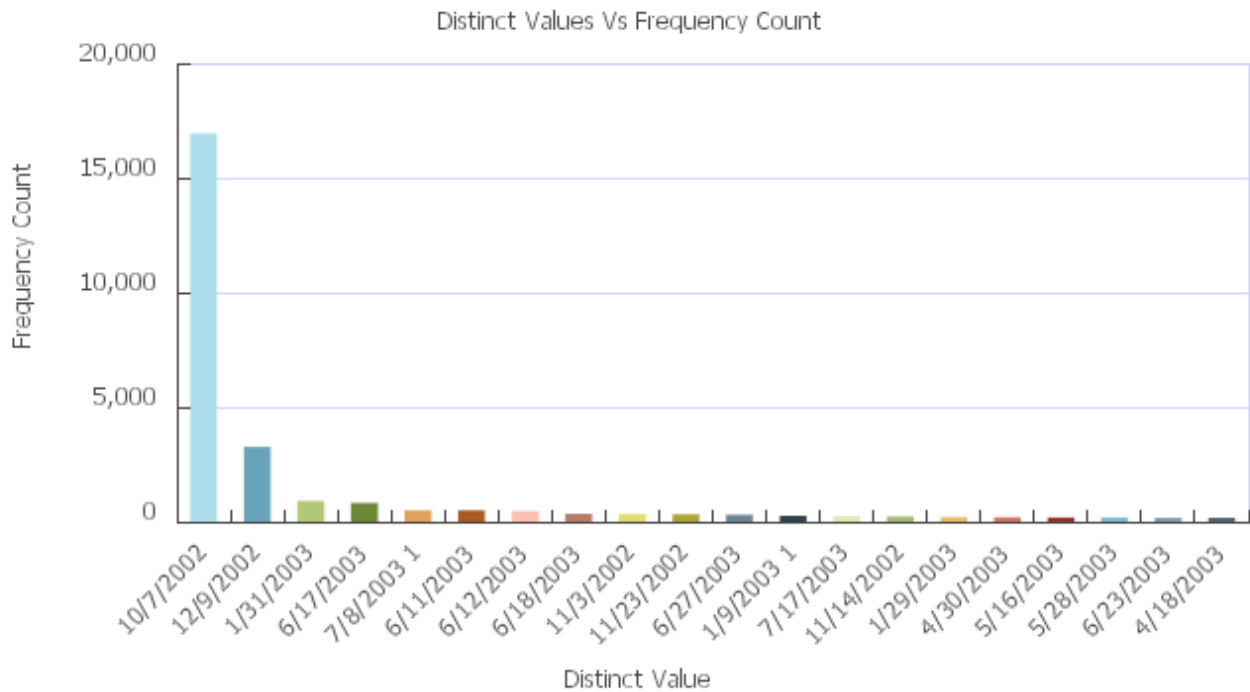
Most Frequent Values



Global Company

Host Name :	IBM-011956133F7
Data Store :	SQLServerE
Data Store Alias :	
Database Name :	dbo
Table Name :	GlobalCoShipTo
Table Alias :	
Column Name :	MOD_DATE

Column Level Summary	
Cardinality Count :	467
Null Count :	0
Actual Row Count :	37712
Total Rows Covered :	32661
%Rows Covered :	86.61





Column Values—Most Frequent Values (Cont.)

Frequency Distribution Data			
Distinct Value	Frequency Count	Frequency %	Cumulative %
10/7/2002	16935	44.91	44.91
12/9/2002	3268	8.67	53.57
1/31/2003	905	2.40	55.97
6/17/2003	820	2.17	58.15
7/8/2003 1	497	1.32	59.46
6/11/2003	496	1.32	60.78
6/12/2003	458	1.21	61.99
6/18/2003	346	.92	62.91
11/3/2002	329	.87	63.78
11/23/2002	318	.84	64.63
6/27/2003	304	.81	65.43
1/9/2003 1	259	.69	66.12
7/17/2003	251	.67	66.79
11/14/2002	231	.61	67.40
1/29/2003	205	.54	67.94
4/30/2003	196	.52	68.46
5/16/2003	185	.49	68.95
5/28/2003	178	.47	69.42
6/23/2003	176	.47	69.89
4/18/2003	170	.45	70.34
4/29/2003	165	.44	70.78
5/30/2003	164	.43	71.21
7/15/2003	163	.43	71.65
11/7/2002	159	.42	72.07
1/6/2003 1	155	.41	72.48
7/16/2003	153	.41	72.88
6/19/2003	152	.40	73.29
3/29/2003	146	.39	73.67
12/10/2002	143	.38	74.05
12/27/2002	143	.38	74.43
1/3/2003 1	140	.37	74.80
3/7/2003 1	140	.37	75.18
6/24/2003	127	.34	75.51
5/21/2003	120	.32	75.83
5/22/2003	118	.31	76.14
5/13/2003	117	.31	76.45
10/10/2002	116	.31	76.76
4/11/2003	116	.31	77.07
5/15/2003	116	.31	77.38
5/27/2003	114	.30	77.68
1/13/2003	113	.30	77.98
1/20/2003	112	.30	78.27
11/27/2002	107	.28	78.56

Column Frequency

Frequency By Frequency



Project:	FK_Test
Report Name :	Frequency By Frequency
Date/ Time Executed :	12/18/2006 04:22:25
Generated By:	admin admin
Customer Description :	Frequency By Frequency
Host Name :	CT-POTTER
Data Store :	myPubs
Data Store Alias :	aaaaaaaaaaaaaaaa
Database Name :	dbo
Table Name :	authors
Table Alias :	
Column Name :	au_lname

Column Level Summary

Cardinality Count:	22
Null Count:	0
Actual Row Count:	23
Frequency Cut Off:	0
Total Rows Covered:	23
%Rows Covered:	100.00

Frequency Distribution Data

Distinct Value	Frequency Count	Frequency%	Cumulative %
Ringer	2	8.70	8.70
Bennet	1	4.35	13.04
Blotchet-Halls	1	4.35	17.39
Carson	1	4.35	21.74
DeFrance	1	4.35	26.09
Dull	1	4.35	30.43
Green	1	4.35	34.78
Greene	1	4.35	39.13
Gringlesby	1	4.35	43.48
Hunter	1	4.35	47.83
Karsen	1	4.35	52.17
Locksley	1	4.35	56.52
MacFeather	1	4.35	60.87
McBadden	1	4.35	65.22
O'Leary	1	4.35	69.57
Panteley	1	4.35	73.91
Smith	1	4.35	78.26
Straight	1	4.35	82.61
Stringer	1	4.35	86.96

Column Frequency

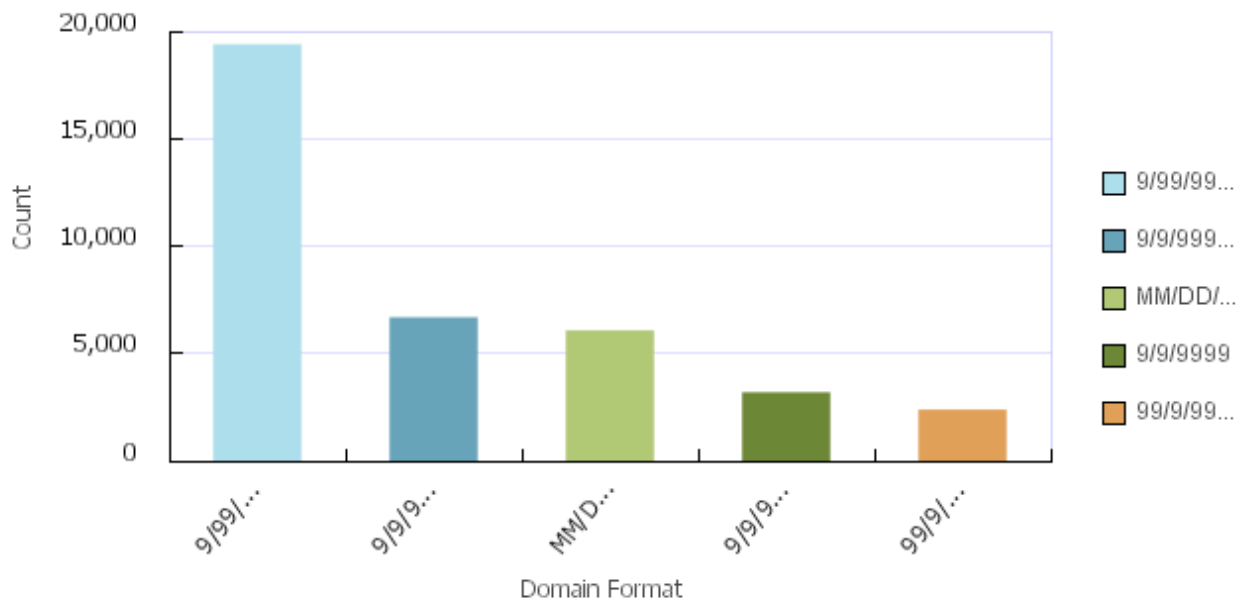
Most Frequent Formats



Column Level Summary

Format Cardinality Count :	5
Value Cardinality Count :	6404
Null Count :	0
Total Actual Row Count :	37712

Most Frequent Format Count



Column Level Details

Format	Count	Total Rows %	Total Rows Cumulative %	Example Values
9/99/9999	19387	51.4080	51.4080	3/13/2002
9/9/9999 9	6681	17.7158	69.1239	4/2/2002 1
MM/DD/YYYY	6069	16.0930	85.2169	12/11/2001
9/9/9999	3188	8.4535	93.6705	7/5/1988
99/9/9999	2387	6.3296	100.0000	10/8/1995
Top Y values :	37,712	100.0000		

Column Analysis—Summary Statistics

Column Analysis

Summary Statistics



Host Name :	IBM-011956133F7
Data Store :	SQLServerE
Data Store Alias:	
Database Name :	dbo
Table Name :	GlobalCoShipTo
Table Name Alias:	

Column Name:	ZIP
Value Summary:	True
Class Code:	

Column Level Content Summary

Total Actual Row Count:	37712
Uniqueness Count:	16,844
Uniqueness %:	44.665
Distinct Value:	20556
Distinct Value %:	54.508
Distinct Format Count:	19
Distinct Format %:	.050
Nullability Flag Chosen:	False
Null Count:	0
Null %:	.000

Column Level Content Summary

Total Actual Row Count:	37712
Uniqueness Count:	16,844
Uniqueness %:	44.665
Distinct Value:	20556
Distinct Value %:	54.508
Distinct Format Count:	19
Distinct Format %:	.050
Nullability Flag Chosen:	False
Null Count:	0
Null %:	.000

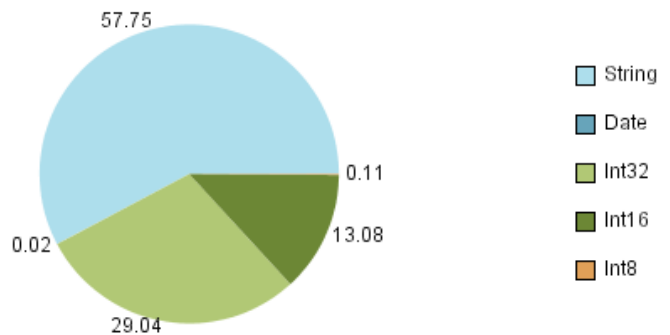
Column Level Inference Structure Summary

Length (Min):	0
Length (Avg):	6.374
Length (Max):	9
Scale (Min):	0
Scale (Max):	0
Precision (Min):	0
Precision (Max):	5

Column Level Structure Definition Summary

Primary Key Flag:	false
Type Code:	String
Length:	255
Precision:	255
Scale:	0

Pie Chart Of Inferred Data Types



Inferred Data Type	Record Count	Record Percent
String	21780	57.754
Date	7	.019
Int32	10950	29.036
Int16	4934	13.083
Int8	41	.109

Column Analysis—Structural Summary

Column Analysis

Structural Summary



Project :	FK_Test
Report Name :	Column Structural Summary
Date/ Time Executed :	12/18/2006 13:24:09
Generated By :	admin admin
Description :	Column Structural Summary

Host Name :	CT-POTTER
Datastore:	myPubs
DataStore Alias:	aaaaaaaaaaaaaaaa
Database Name :	dbo
Table Name :	authors
Table Alias:	

Column Name:	au_lname
Total Records:	23
Content:	90.000
Content Domain:	10.000

Column Level Structure Definition Summary

Primary key Flag:	false
# Of Inferred Data types:	1
Length:	40
Precision:	40
Scale:	0

Column Level Content Summary

Unique Values:	21
Uniqueness %:	91.304
Distinct Format Count:	10
Distinct Format %:	43.478
Null %:	.000

Column Level Domain Summary

Incomplete %:	Null %
Invalid %:	.000
Minimum Value:	Bennet
Maximum Value:	del Castillo


Column Domain Values

Domain Value Profile Report

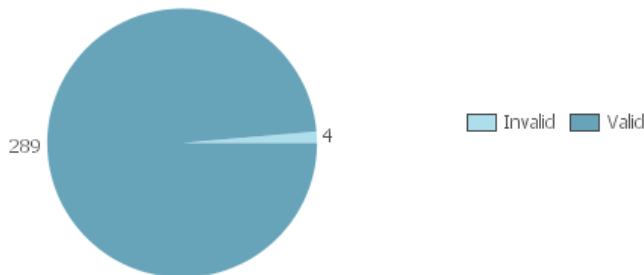
Product Module :	WebSphere Information Analyzer
Project :	GlobalCoIntegration
Report Name :	Domain Value Profile
Date/ Time Executed :	03/01/2007 14:09:48
Generated By :	admin admin
Customer Description :	Domain Value Profile
Host Name :	IBM-011956133F7
Data Store :	SQLServerE
Data Store Alias :	
Database Name :	dbo
Table Name :	GlobalCoShipTo
Table Alias :	
Column Name :	CNTY_NAME
Data Class :	
Sort Order :	Descending by Count of Value

Column Level Detail

Domain Indicator	Value	Count of Value	Percent of Total
Default	[SPACES]	29835	79.11%
Valid	G	2306	6.11%
Valid	F	2069	5.49%
Valid	A	1918	5.09%
Valid	D	1552	4.12%
Default	[EMPTY]	10	0.03%
Valid	P	10	0.03%
Valid	S	9	0.02%
Valid	B	2	0.01%
Valid	I	1	0%

Column Domain Values		 Global Company
Domain Analysis Detail		
Product Module :	WebSphere Information Analyzer	
Project :	GlobalCoIntegration	
Report Name :	Domain Analysis Detail	
Date/ Time Executed :	03/01/2007 13:51:51	
Generated By :	admin admin	
Customer Description :	GlobalCo ShipTo - Domain Analysis	
Host Name :	IBM-011956133F7	
Data Store :	SQLServerE	
Data Store Alias :		
Database Name :	dbo	
Table Name :	GlobalCoShipTo	
Table Alias :		
Column Name :	CFA_AREA_CODE	
Data Class :		
Sort Order :	Descending by Count of Value	
Excluded Indicator(s) :	Valid	

Summary of Domain Indicator Percents for the Column



Domain Value Profile Detail

Value	Count of Value	Percent of Total	Domain Indicator	Mapped Value	Value Definition
XXX	2	0.005%	Invalid		
.	1	0.003%	Invalid		
SEE	1	0.003%	Invalid		
TO	1	0.003%	Invalid		

CFA Phone data contains embedded notes

Type : Action Status : Opened Created : 03/01/07 Modified : 03/01/07

Subject :

Check out this account:
SHIP_TO_NUM GC86226

Embedded data:
SEE FAX NOTE

Column Analysis—Column Properties (by table)

Column Properties

Definition Summary

Product Module :	WebSphere Information Analyzer
Project :	RetailCoAssessment
Report Name :	RetailCo - Definition Summary
Date/ Time Executed :	03/01/07 12:33:47
Generated By :	admin admin
Customer Description :	RetailCo - Product Master Summary
Host Name :	IBM-011956133F7
Data Store :	SQLServerC
Data Store Alias :	
Database Name :	dbo
Table Name :	Product_Master
Table Alias :	

Column Level Summary						
Column Name	Defined Data Type	Defined Length	Defined Precision	Defined Scale	Defined Nullity	Defined Key Flag
Cost	Decimal	19	19	4	True	none
Description	String	500	none	0	True	none
Product_Family	String	50	none	0	True	none
Product_Family_Plan	Decimal	19	19	4	True	none
Product_Id	Int32	4	4	0	False	none
Product_Name	String	50	none	0	True	none
Product_Plan	Decimal	19	19	4	True	none
Supplier_Id	Int32	4	4	0	True	none
UOM	String	25	none	0	True	none
Vendor	String	50	none	0	True	none
Wholesale_Price	Decimal	19	19	4	True	none

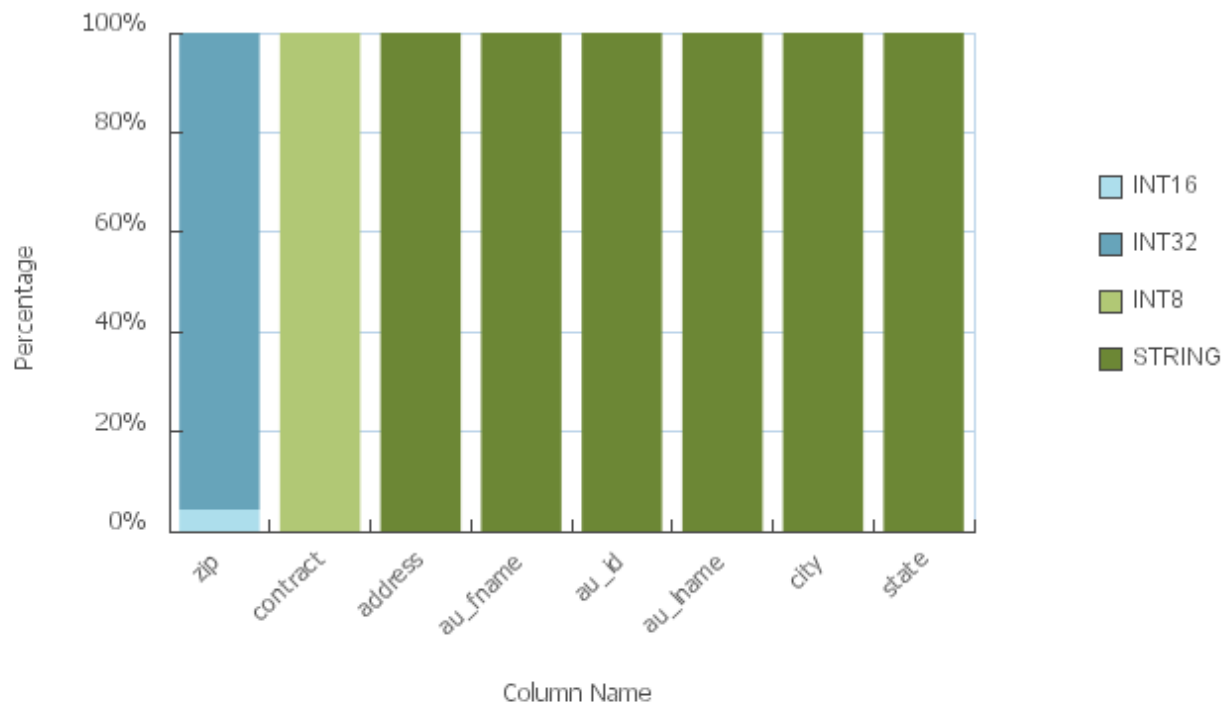
Column Analysis—Column Data Types (by table)

Column Properties	
Column All Data Types	
Project :	FK_Test
Report Name :	Column All Data Types
Date/ Time Executed :	12/18/06 13:16:55
Generated By :	admin admin
Customer Description :	Column All Data Types
Host Name :	CT-POTTER
Data Store :	myPubs
Data Store Alias :	aaaaaaaaaaaaaaaa
Database Name :	dbo
Table Name :	authors
Table Alias :	



Column Summary

Candidate Data Type Distribution



Primary Key Analysis—Candidate Duplicate Exceptions



Primary Key Analysis

Candidate Duplicate Exceptions

Project :	Test_FK_Reports
Report Name :	Candidate Duplicate Exceptions
Date/ Time Executed :	12/18/06 14:01:49
Generated By :	admin admin
Customer Description :	Candidate Duplicate Exceptions
Host Name :	CT-POTTER
Data Store :	myPubs
Data Store Alias :	aaaaaaaaaaaaaaaa
Database Name :	dbo
Table Name :	authors
Table Alias :	

Table Level Summary

Total Rows :	23
--------------	----

Table Level Details

Primary Key Columns	PK Duplicate Value	# of Duplicates	% Duplicates
address			
	3410 Blonde St.	2	8.696%
	67 Seventh Av.	2	8.696%
au_fname			
	*003	*003	*003
au_id			
	*003	*003	*003
au_lname			
	Ringer	2	8.696%
city			
	Oakland	5	21.739%
	Berkeley	2	8.696%
	Palo Alto	2	8.696%
	Salt Lake City	2	8.696%
contract			
	1	19	82.609%
	0	4	17.391%
phone			
	415 836-7128	2	8.696%
	801 826-0752	2	8.696%
state			

*003 - No duplicate values found

Column Analysis—Primary Key Chosen

Primary Key Analysis

Defined Primary Key - Summary



Project :	Test_FK_Reports
Report Name :	Defined Primary Key - Summary
Date/ Time Executed :	12/18/06 14:04:37
Generated By :	admin admin
Customer Description :	Defined Primary Key - Summary
Host Name :	CT-POTTER
Data Store :	myPubs
Data Store Alias :	aaaaaaaaaaaaaaaa
Database Name :	dbo
Table Name :	authors
Table Alias :	
PK Type Flag :	D


Defined Primary Key Summary

Primary Key Column(s)	PK Issue Flag	Length		Null Values		Default Values		Duplicate Values		Uniqueness	Distinct Values	Total Rows
		Defined	Inferred	Total Rows	% of Total Rows	Total Rows	% of Total Rows	Total Rows	% of Total Rows	% of Total Rows	% of Total Rows	
au_id	Y	11	11	0	.000%	0	0.000%	0	0.000%	100.000%	100.000%	23

Column Analysis—Primary Key Candidates

Primary Key Analysis

Defined & Candidate Summary



Project :	Test_FK_Reports
Report Name :	Defined and Candidate Summary
Date/ Time Executed :	12/18/06 14:06:07
Generated By :	admin admin
Customer Description :	Defined and Candidate Summary
Host Name :	CT-POTTER
Data Store :	myPubs
Data Store Alias :	aaaaaaaaaaaaaaaa
Database Name :	dbo
Table Name :	authors
Table Alias :	

Defined Primary Key Summary

Primary Key Column(s)	PK Issue Flag	PK Type Flag	Length		Null Values		Default Values		Duplicate Values		Uniqueness %	
			Defined	Inferred	Total	%	Total	%	Total	%	Distinct Values %	Total Rows
au_fname	Y	S	20		0		0		0		100.000%	
				11	0	0%	0	0%	0	0%	100%	23
au_id	Y	D	11		0		0		0		100.000%	
				11	0	0%	0	0%	0	0%	100%	23
au_fname	N	C	20		0		0	None	None	None	None	None
				11	0	0%	0	None	None	None	None	None

IBM WebSphere Information Analyzer Profiling and Analysis Reports

For more information

For more information about WebSphere Information Analyzer, call your IBM representative or visit us online at ibm.com/software/data/integration/information_analyzer



© Copyright IBM Corporation 2007

IBM Software Group
Route 100
Somers, NY 10589

Produced in the United States
March 2007
All Rights Reserved

IBM, the IBM logo and WebSphere are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Other company, product or service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice.

All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only.

***TAKE BACK CONTROL WITH* Information Management**