Scott Refugio
BANA 290
Assignment 2

Questions to answer:
• How many tables does it have?
• How many of those sentences and tables were identified by IBM Compare & Comply?
• Did Compare & Comply identify correct number of rows and columns for each table?
• Did Compare & Comply identify row headers, column headers, table headers?
• Report it in the format provided, so we can pool across the class.
• How complex are the tables in your report?
• What was precision and recall?

Company	# of Tables		Precision	Recall	F1	Complexity
	detected	actual				
America's Car-Mart	28	22	100%	78.5%	0.879	Tables included a balance sheet (maturity level 3), a columnless table (maturity level 2), a table with a column header with subheaders (maturity level 4), and a table with empty rows (maturity level 3). Compare & Comply had detected a number of objects as tables, but these objects can be considered tables with 1 row and two columns. I did not interpret these as tables in the analysis.
BJ's Restaurants Inc.	15	14	100%	93.3%	0.965	Tables included a simple table (maturity level 1), a statement of income (maturity level 3), a short summary table that is near another table (maturity level 2), and a consolidated statement of income (maturity level 3). Compare & Comply's F1 score was relatively high. Compared to the other documents, this document had fairly simple tables in terms of structure of simple datatypes.
Mesa Laboratories Inc.	23	32	71.9%	100%	0.837	Tables included an income statement that spanned multiple pages (maturity level 4), tables with semi-structured headings (maturity level 4), and tables of various formats (maturity level 4). Compare & Comply had the most difficulty with this document, partly attributed to tables spanning multiple pages and headers containing subheaders.
Quality Systems Inc.	4	5	80.0%	100%	0.888	Tables included a table of contents (maturity level 1), a table without column headers (maturity level 2), a table with empty cell bodies (maturity level 3), and a table where row headers were located in the same column as the row (maturity level 4). Compare & Comply failed to detect a maturity level 5 table that was fairly unstructured, with no clear borders or column/row lines.

Quality Systems Inc.

Table		1	2	3	4	Precision	Recall	F1
Row Headers	classified	2	14	8	0	100.0%	68.5%	0.813

	actual	2	14	8	11			
Column Headers	classified	0	0	0	0	0%	0%	NaN
	actual	1	1	2	3			
Table Header	classified	0	1	0	0	0%	0%	NaN
	actual	0	0	1	0			
Rows	classified	2	14	7	15	86.8%	97.1%	0.917
	actual	2	14	8	10			
Columns	classified	4	1	1	3	100.0%	90.0%	0.947
	actual	5	1	1	3			

Comments: Row headers were difficult for Compare & Comply to classify for the fourth table as the headers were located in the same column as the row. Rows were difficult in the fourth table as there were rows with various heights.

Mesa Labs Inc.

Table		1	2	3	4	Precision	Recall	F1
Row Headers	classified	36	3	3	12	83.3%	100%	0.909
	actual	35	3	3	4			
Column Headers	classified	9	5	8	0	36.4%	100%	0.534
	actual	2	2	4	0			
Table Header	classified	0	0	0	0	0%	0%	NaN
	actual	1	1	0	0			
Rows	classified	38	4	4	11	82.5%	100%	0.904
	actual	36	3	4	4			
Columns	classified	4	2	3	2	72.7%%	88.9%	0.799
	actual	2	2	4	1			

Comments: Column headers were difficult for Compare & Comply for all tables as there were hierarchies of column subheaders for most tables.

Scott Refugio
BANA 290
Assignment 2

BJ's Restaurants Inc.

Table		1	2	3	4	Precision	Recall	F1
Row Headers	classified	3	20	3	0	96.2%	75.8%	0.848
	actual	3	19	3	8			
Column Headers	classified	0	16	3	42	9.1%	100%	0.166
	actual	0	4	2	5			
Table Header	classified	0	0	0	0	0%	0%	NaN
	actual	0	0	0	0			
Rows	classified	2	23	3	14	76.2%	96.9%	0.853
	actual	3	19	3	8			
Columns	classified	1	4	2	15	54.5%	100.0%	0.705
	actual	1	4	2	5			

Comments: Column headers were difficult for Compare & Comply for all tables as there were hierarchies of column subheaders for most tables.

America's Car-Mart Inc.

Table		1	2	3	4	Precision	Recall	F1
Row Headers	classified	0	0	0	4	11.1%	100%	0.199
	actual	23	3	6	4			
Column Headers	classified	8	0	6	3	35.3%	100%	0.522
	actual	2	0	2	2			
Table Header	classified	0	0	1	0	0%	0%	NaN
	actual	0	0	0	0			
Rows	classified	27	2	7	4	85.0%	97.1%	0.906
	actual	23	3	5	4			
Columns	classified	2	2	2	2	87.5%	100%	0.933
	actual	2	1	2	2			

Scott Refugio
BANA 290
Assignment 2

Comments: Row headers were difficult for Compare & Comply, possibly due to the lengthy text in each row header. Column headers were difficult for Compare & Comply for all tables as there were hierarchies of column subheaders for most tables.