

IBM-SPSS

Comprehensive Data Mining Projects

BY: VISHAL K ANANT

INDEX

<u>S.NO</u>	<u>TITLE</u>
1.	PROJECT-1 Creating a Data-Mining Project
2.	PROJECT-2 Collecting Initial Data
3.	PROJECT-3 Understanding your Data
4.	PROJECT-4 Setting the Unit of Analysis
5.	PROJECT-5 Integrating Data
6.	PROJECT-6 Deriving and Reclassifying the Fields

PROJECT-1

Creating a Data-Mining Project

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models Stream1

Fields - 12 Records - 30000

Table ACME customer and rf...

CRISP-DM Classes

(unsaved project) Business Understanding Data Understanding Data Preparation Modeling Evaluation Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 134MB / 150MB

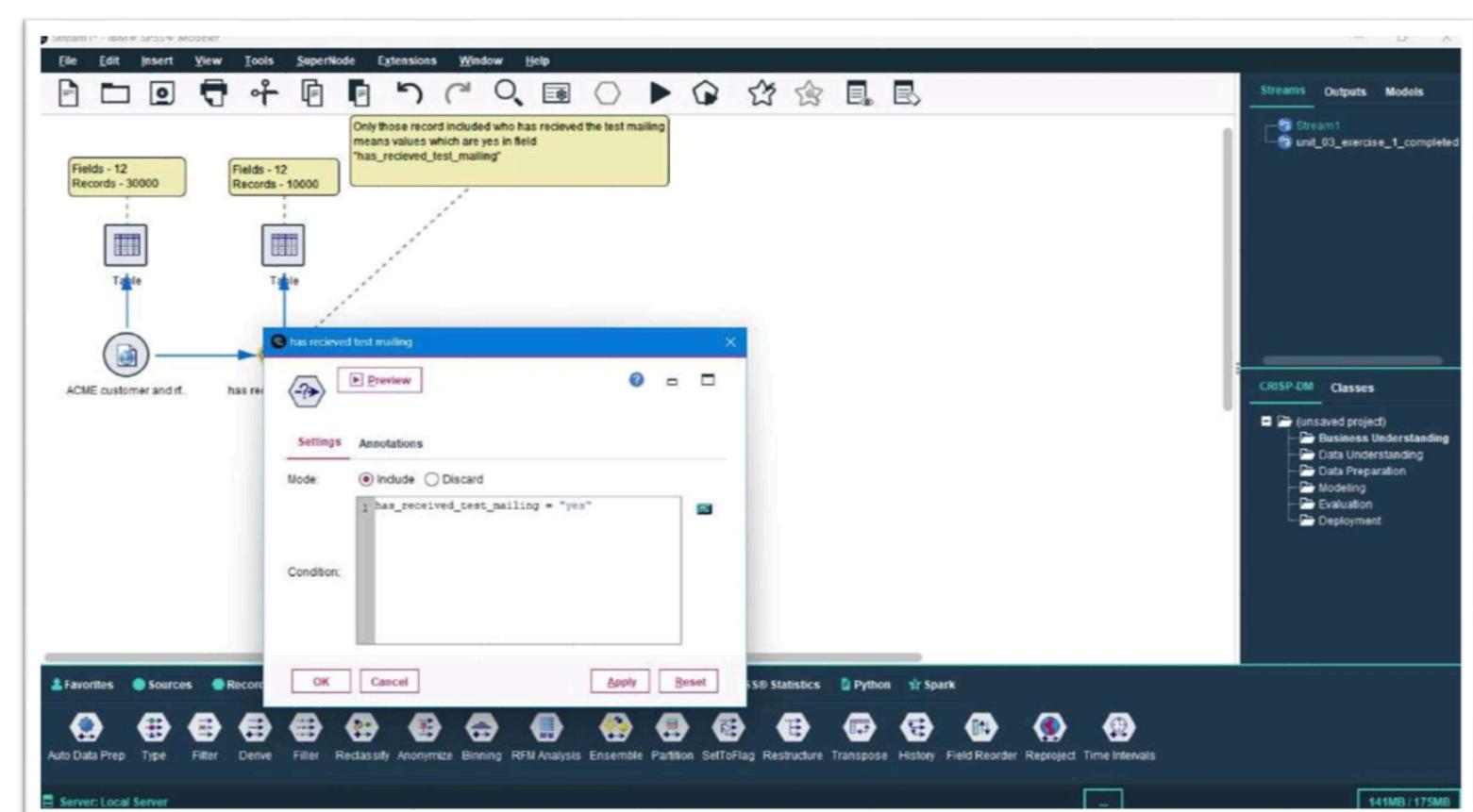
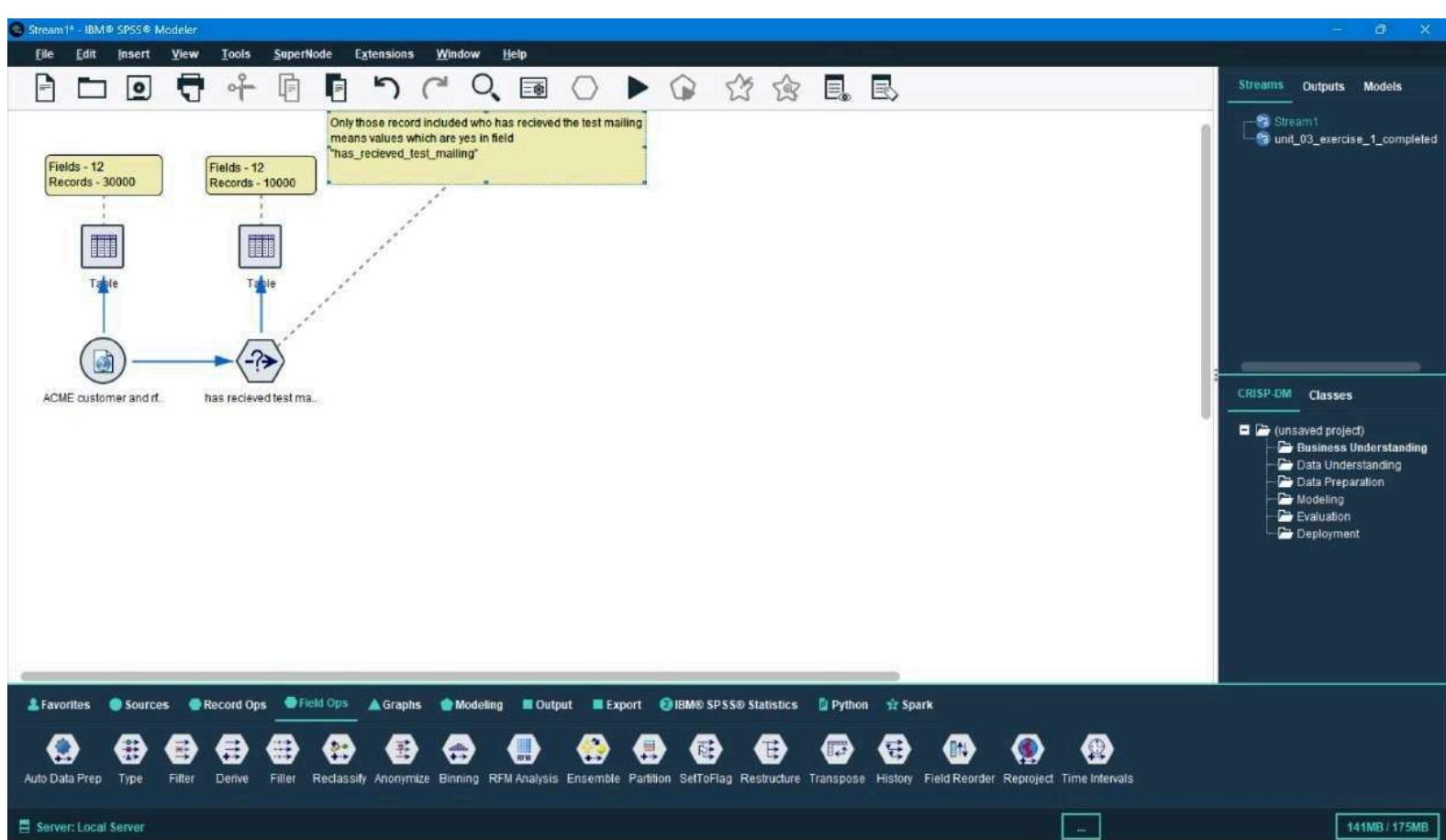
Table (12 fields, 30,000 records) #3

File Edit Generate

Annotations

	customer_id	gender	email_address	postal_code	monetary_value_01_01_2011	frequency_01_01_2011	recency_01_01_2011	has_received_test_mailing	response_to_test_mailing_02_01_2011	orderdate	number_of_days_between_mailing
1	723.000	male	name7502@mail.fr	181B0	2 medium	3 high	2 medium	yes	F	\$null\$	
2	724.000	female	name25485@wwmail.org	1132DG	1 low	3 high	1 low	yes	F	\$null\$	
3	725.000	male	name15543@wwmail.de	1803YT	3 high	1 low	1 low	yes	F	\$null\$	
4	726.000	male	name28336@zigzag.be	1205WR	3 high	1 low	3 high	yes	F	\$null\$	
5	727.000	female	name5354@tnet.jp	1711ON	1 low	3 high	1 low	yes	F	\$null\$	
6	728.000	female	name20537@wwmail.es	1055FG	2 medium	3 high	1 low	yes	T	2011-02-04	
7	729.000	female	name20536@wwmail.es	1254MR	1 low	3 high	1 low	yes	F	\$null\$	
8	730.000	female	name10414@tnet.linc	1723DG	2 medium	3 high	1 low	yes	F	\$null\$	
9	731.000	male	name23372@wwmail.inc	1713AQ	3 high	2 medium	1 low	yes	F	\$null\$	
10	732.000	male	name20635@wwmail.es	1264EC	3 high	2 medium	3 high	yes	T	2011-02-15	
11	733.000	female	name5356@tnet.jp	1648BT	3 high	2 medium	1 low	yes	F	\$null\$	
12	734.000	female	name17582@wwmail.de	1285XV	3 high	1 low	3 high	yes	F	\$null\$	
13	735.000	female	name6388@tnet.fr	1282NB	1 low	2 medium	2 medium	yes	F	\$null\$	
14	736.000	male	name10409@tnet.linc	1799T	3 high	2 medium	1 low	yes	F	\$null\$	
15	737.000	female	name13849@tnet.uk	1802UO	2 medium	3 high	1 low	yes	F	\$null\$	
16	738.000	male	name25473@wwmail.org	1971NK	1 low	3 high	1 low	yes	F	\$null\$	
17	739.000	male	name13848@tnet.uk	1361RL	2 medium	3 high	1 low	yes	F	\$null\$	
18	740.000	female	name23366@wwmail.in	1164VN	3 high	2 medium	1 low	yes	F	\$null\$	
19	741.000	female	name3188@molbe.cat	1767YN	3 high	1 low	1 low	yes	F	\$null\$	
20	742.000	male	name1606@lomejor.es	1681HP	1 low	3 high	1 low	yes	F	\$null\$	
21	743.000	female	name29056@zigzag.cat	1287PI	3 high	2 medium	1 low	yes	F	\$null\$	
22	744.000	male	name28323@zigzag.be	1023AR	2 medium	3 high	1 low	yes	T	2011-02-28	
23	745.000	female	name20630@wwmail.es	1737ZI	3 high	2 medium	1 low	yes	F	\$null\$	
24	746.000	male	name13847@tnet.uk	1625TZ	1 low	3 high	1 low	yes	F	\$null\$	
25	747.000	male	name10401@tnet.linc	1684YE	1 low	3 high	2 medium	yes	F	\$null\$	
26	748.000	female	name13846@tnet.uk	1508DU	2 medium	2 medium	1 low	yes	F	\$null\$	
27	749.000	female	name17579@wwmail.de	1143VC	1 low	3 high	2 medium	yes	F	\$null\$	
28	750.000	female	name26094@wwmail.uk	1228H	3 high	1 low	2 medium	yes	F	\$null\$	
29	751.000	female	name6366@tnet.fr	1567CX	3 high	2 medium	3 high	yes	T	2011-02-17	
30	752.000	male	name1609@lomejor.es	1315FI	3 high	2 medium	1 low	yes	F	\$null\$	
31	753.000	female	name29721@zigzag.com	1753OI	1 low	3 high	1 low	yes	F	\$null\$	
32	754.000	female	name29850@zigzag.es	1229DC	1 low	3 high	3 high	yes	F	\$null\$	
33	755.000	female	name17578@wwmail.de	1038LJ	2 medium	3 high	1 low	yes	T	2011-02-27	
34	756.000	female	name29046@zigzag.cat	1002VW	3 high	2 medium	1 low	yes	T	2011-02-22	
35	757.000	female	name10389@tnet.linc	1979KZ	3 high	3 high	1 low	yes	F	\$null\$	
36	758.000	male	name6354@tnet.fr	1784IE	3 high	1 low	3 high	yes	F	\$null\$	
37	759.000	male	name3172@molbe.cat	1183FC	3 high	2 medium	1 low	yes	F	\$null\$	
38	760.000	male	name43@mail.de	1492WX	1 low	3 high	1 low	yes	F	\$null\$	
39	761.000	female	name5357@tnet.jp	1569YY	3 high	2 medium	1 low	yes	F	\$null\$	
40	762.000	male	name256@mail.linc	1558OG	1 low	3 high	1 low	yes	F	\$null\$	
41	763.000	male	name25457@wwmail.org	1598SB	3 high	3 high	3 high	yes	T	2011-02-26	
42	764.000	female	name1616@lomejor.es	1972BQ	2 medium	3 high	1 low	yes	F	\$null\$	
43	765.000	male	name6359@tnet.in	1100VN	2 medium	2 medium	1 low	yes	F	\$null\$	

OK



Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Fields - 12 Records - 30000 Fields - 12 Records - 10000

Only those record included who has received the test mailing means values which are yes in field "has_received_test_mailing"

Table Table

ACME customer and rf...

has received test mailing

Settings Annotations

Name: Auto Custom has received test mailing

Tooltip text:

Keywords:

Created on 23 December, 2022 3:33 PM by dell. ID: id7HKV1MTHDG

This object has not been saved.

OK Cancel Apply Reset

Favorites Sources Records

Auto Data Prep Type Filter Derive Filler Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproj. Time Intervals

CRISP-DM Classes

(unsaved project) Business Understanding Data Understanding Data Preparation Modeling Evaluation Deployment

Server: Local Server 141MB / 175MB

Table (12 fields, 10,000 records) #5

Annotations

	customer_id	gender	email_address	postal_code	monetary_value_01_01_2011	frequency_01_01_2011	recency_01_01_2011	has_received_test_mailing	response_to_test_mailing_02_01_2011	orderdate	number_of_days_between_mailing
1	723.000	male	name7502@tnet.fr	1818BQ	2 medium	3 high	2 medium	yes	F	\$null\$	
2	724.000	female	name25485@wmail.org	1132DG	1 low	3 high	1 low	yes	F	\$null\$	
3	725.000	male	name15543@wmail.de	1803YT	3 high	1 low	1 low	yes	F	\$null\$	
4	726.000	male	name28335@zigzag.be	1205WR	3 high	1 low	3 high	yes	F	\$null\$	
5	727.000	female	name5354@tnet.jp	1711ON	1 low	3 high	1 low	yes	F	\$null\$	
6	728.000	female	name20637@wmail.es	1056FG	2 medium	3 high	1 low	yes	T	2011-02-04	
7	729.000	female	name20536@wmail.es	1254IR	1 low	3 high	1 low	yes	F	\$null\$	
8	730.000	female	name10414@tnet.lne	1723DG	2 medium	3 high	1 low	yes	F	\$null\$	
9	731.000	male	name23372@wmail.inc	1713AQ	3 high	2 medium	1 low	yes	F	\$null\$	
10	732.000	male	name20635@wmail.es	1264EC	3 high	2 medium	3 high	yes	T	2011-02-15	
11	733.000	female	name5356@tnet.jp	1648BT	3 high	2 medium	1 low	yes	F	\$null\$	
12	734.000	female	name17582@wmail.de	1285XV	3 high	1 low	3 high	yes	F	\$null\$	
13	735.000	female	name6388@tnet.fr	1282NB	1 low	2 medium	2 medium	yes	F	\$null\$	
14	736.000	male	name10409@tnet.inc	1799IT	3 high	2 medium	1 low	yes	F	\$null\$	
15	737.000	female	name13849@tnet.uk	1802UO	2 medium	3 high	1 low	yes	F	\$null\$	
16	738.000	male	name25473@wmail.org	1971NK	1 low	3 high	1 low	yes	F	\$null\$	
17	739.000	male	name13848@tnet.uk	1361RL	2 medium	3 high	1 low	yes	F	\$null\$	
18	740.000	female	name23366@wmail.inc	1164VN	3 high	2 medium	1 low	yes	F	\$null\$	
19	741.000	female	name3188@molbe.cat	1767YN	3 high	1 low	1 low	yes	F	\$null\$	
20	742.000	male	name1606@lomejor.es	1681HP	1 low	3 high	1 low	yes	F	\$null\$	
21	743.000	female	name29050@zigzag.ca	1287PI	3 high	2 medium	1 low	yes	F	\$null\$	
22	744.000	male	name28323@zigzag.be	1023AR	2 medium	3 high	1 low	yes	T	2011-02-28	
23	745.000	female	name20630@wmail.es	1737ZI	3 high	2 medium	1 low	yes	F	\$null\$	
24	746.000	male	name13847@tnet.uk	1625TZ	1 low	3 high	1 low	yes	F	\$null\$	
25	747.000	male	name10401@tnet.inc	1684YE	1 low	3 high	2 medium	yes	F	\$null\$	
26	748.000	female	name13846@tnet.uk	1508DU	2 medium	2 medium	1 low	yes	F	\$null\$	
27	749.000	female	name17579@wmail.de	1143VC	1 low	3 high	2 medium	yes	F	\$null\$	
28	750.000	female	name26894@wmail.uk	1228IK	3 high	1 low	2 medium	yes	F	\$null\$	
29	751.000	female	name5366@tnet.fr	1567CX	3 high	2 medium	3 high	yes	T	2011-02-17	
30	752.000	male	name1609@lomejor.es	1315FI	3 high	2 medium	1 low	yes	F	\$null\$	
31	753.000	female	name29727@zigzag.com	1753OI	1 low	3 high	1 low	yes	F	\$null\$	
32	754.000	female	name29856@zigzag.es	1229DC	1 low	3 high	3 high	yes	F	\$null\$	
33	755.000	female	name17578@wmail.de	1036LJ	2 medium	3 high	1 low	yes	T	2011-02-27	
34	756.000	female	name29046@zigzag.ca	1002VV	3 high	2 medium	1 low	yes	T	2011-02-22	
35	757.000	female	name10389@tnet.inc	1979KZ	3 high	3 high	1 low	yes	F	\$null\$	
36	758.000	male	name6354@tnet.fr	1784IE	3 high	1 low	3 high	yes	F	\$null\$	
37	759.000	male	name3172@molbe.cat	1183FC	3 high	2 medium	1 low	yes	F	\$null\$	
38	760.000	male	name43@mail.de	1492WX	1 low	3 high	1 low	yes	F	\$null\$	
39	761.000	female	name5357@tnet.jp	1569YY	3 high	2 medium	1 low	yes	F	\$null\$	
40	762.000	male	name255@hmail.inc	15580G	1 low	3 high	1 low	yes	F	\$null\$	
41	763.000	male	name25457@wmail.org	1588SB	3 high	3 high	3 high	yes	T	2011-02-26	
42	764.000	female	name1616@lomejor.es	1972BQ	2 medium	3 high	1 low	yes	F	\$null\$	
43	765.000	male	name5359@tnet.jp	1109D	2 medium	2 medium	1 low	yes	F	\$null\$	

OK

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

```

graph LR
    A[Fields - 12  
Records - 30000] --> B[Table]
    B --> C{?}
    D[Fields - 12  
Records - 10000] --> E[Table]
    E --> F{?}
    C --> G{Type}
    F --> G
    
```

ACME customer and rf... has received test ma...

Only those record included who has received the test mailing means values which are yes in field "has_received_test_mailing"

Setting Roles

Input - gender, monetary_value_01_01_2011, frequency_01_01_2011, recency_01_01_2011

Target - response_to_test_mailing_02_01_2011

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

All Supervised Association Segmentation KNN Isotonic-AS Extension Model XGBoost Linear XGBoost Tree CHAID Quest Gaussian Mixture KDE Modeling STP Association Rules XGBoost-AS Random Forest K-Means-AS HDBSCAN MultiLayerPerceptron-AS One-Class SVM

Server: Local Server 143MB / 175MB

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Only those record included who has received the test mailing means values which are yes in field "has_received_test_mailing"

Type

Field	Measurement	Values	Check	Role
gender	Nominal	male, female	None	Input
email_address	Typeless		None	None
postal_code	Typeless		None	None
monetary_value_01_01_2011	Nominal	"1 low", "3 high"	None	Input
frequency_01_01_2011	Nominal	"1 low", "3 high"	None	Input
recency_01_01_2011	Nominal	"1 low", "3 high"	None	Input
has_received_test_mailing	Nominal	no, yes	None	Input
response_to_test_mailing_02_01_2011	Nominal	F, T	None	Target

View current fields View unused field settings

OK Cancel Apply Reset

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

All Supervised Association Segmentation KNN Isotonic-AS Extension Model XGBoost Linear XGBoost Tree CHAID Quest Gaussian Mixture KDE Modeling STP Association Rules XGBoost-AS Random Forest K-Means-AS HDBSCAN MultiLayerPerceptron-AS One-Class SVM

Server: Local Server 144MB / 175MB

Stream1 - IBM SPSS Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

All Supervised Association Segmentation KNN Isotonic-AS Extension Model XGBoost Linear XGBoost Tree CHAID Quest Gaussian Mixture KDE Modeling STP Association Rules XGBoost-AS Random Forest K-Means-AS HDBSCAN MultiLayerPerceptron-AS One-Class SVM

Server: Local Server 161MB / 175MB

Stream1 - IBM SPSS Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

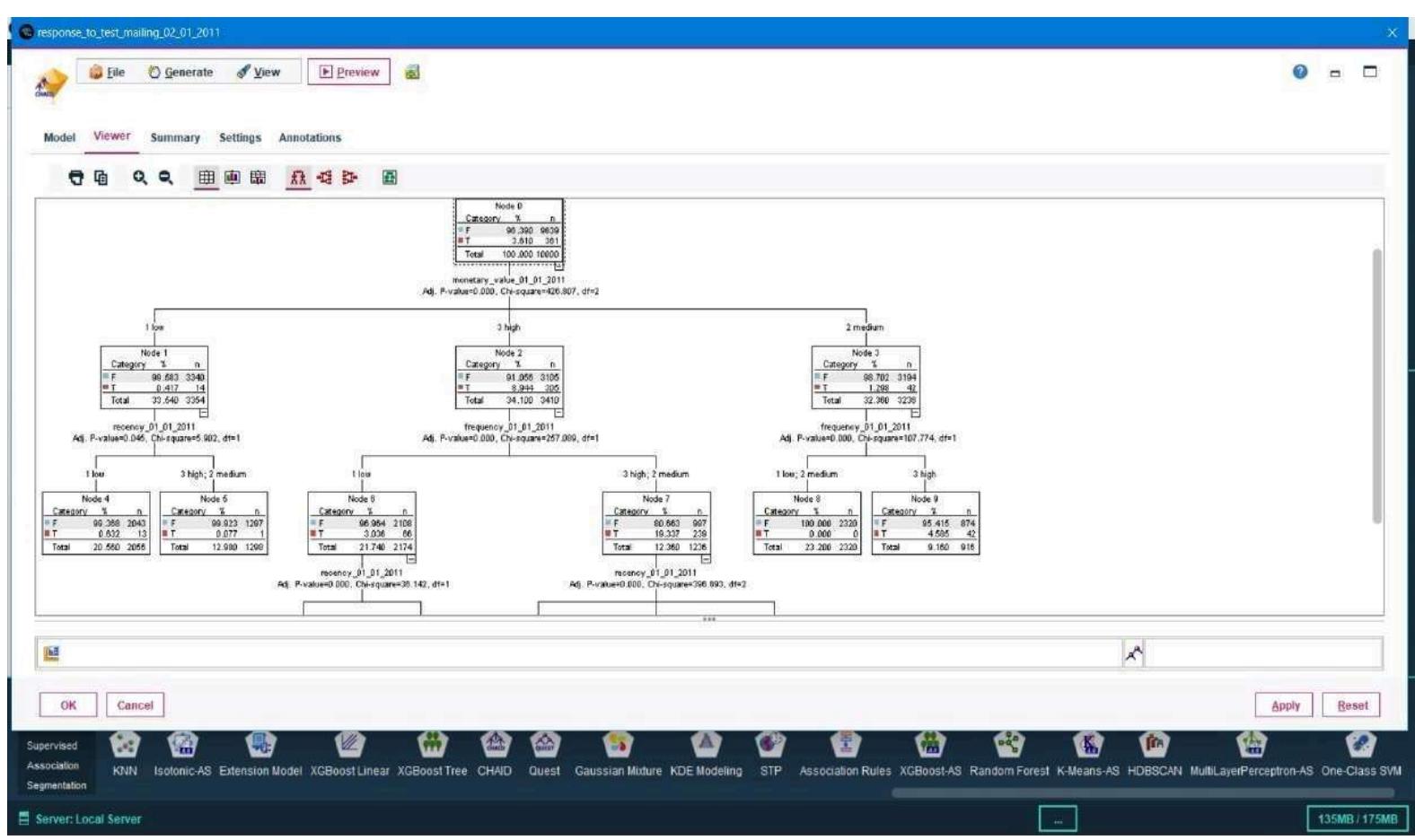
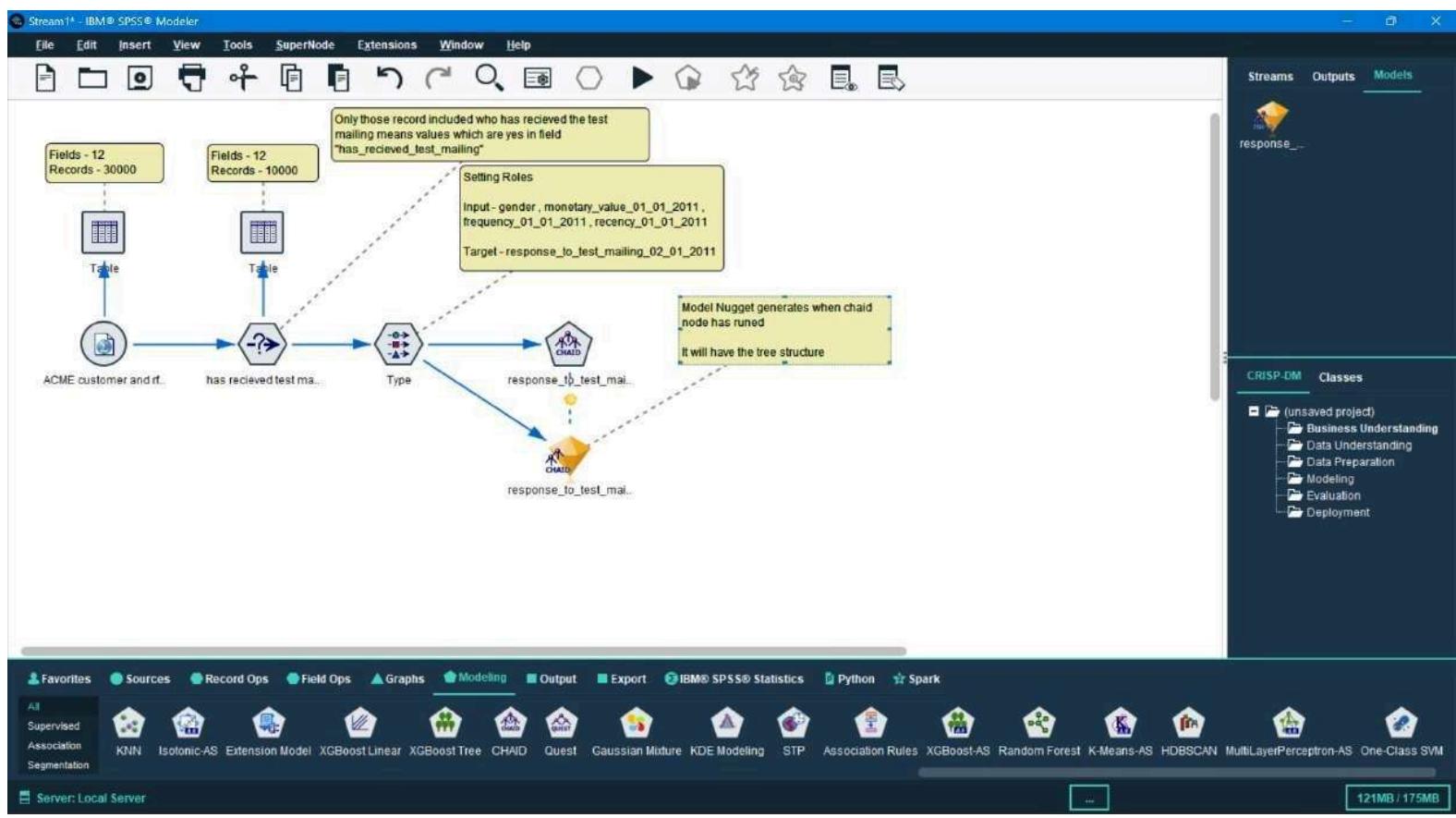
CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

All Supervised Association Segmentation KNN Isotonic-AS Extension Model XGBoost Linear XGBoost Tree CHAID Quest Gaussian Mixture KDE Modeling STP Association Rules XGBoost-AS Random Forest K-Means-AS HDBSCAN MultiLayerPerceptron-AS One-Class SVM

Server: Local Server 161MB / 175MB



Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

```

graph LR
    A[Fields - 12  
Records - 30000] --> B[Table]
    B --> C{?}
    C --> D{Type}
    D --> E[CHAID]
    E --> F[CHAID]
    F --> G[Table]
    
```

Annotations:

- Only those records included who has received the test mailing means values which are yes in field "has_received_test_mailing"
- Setting Roles
Input - gender, monetary_value_01_01_2011, frequency_01_01_2011, recency_01_01_2011
Target - response_to_test_mailing_02_01_2011
- Model Nugget generates when chaid node has runned
It will have the tree structure
- Fields - 14
2 extra fields added
- \$R-response - gives predicted response
- \$RC-response - gives the confidence for prediction

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

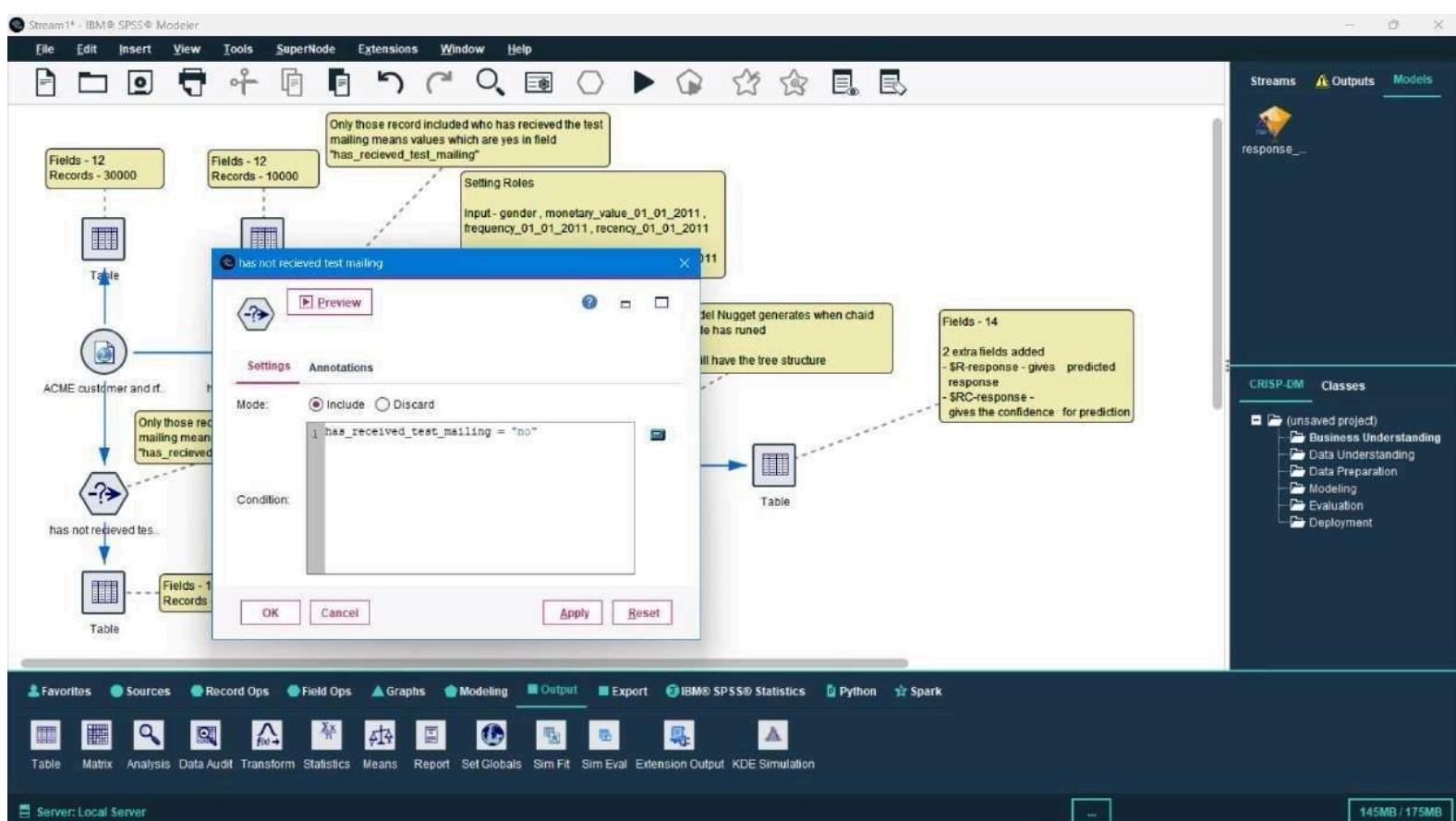
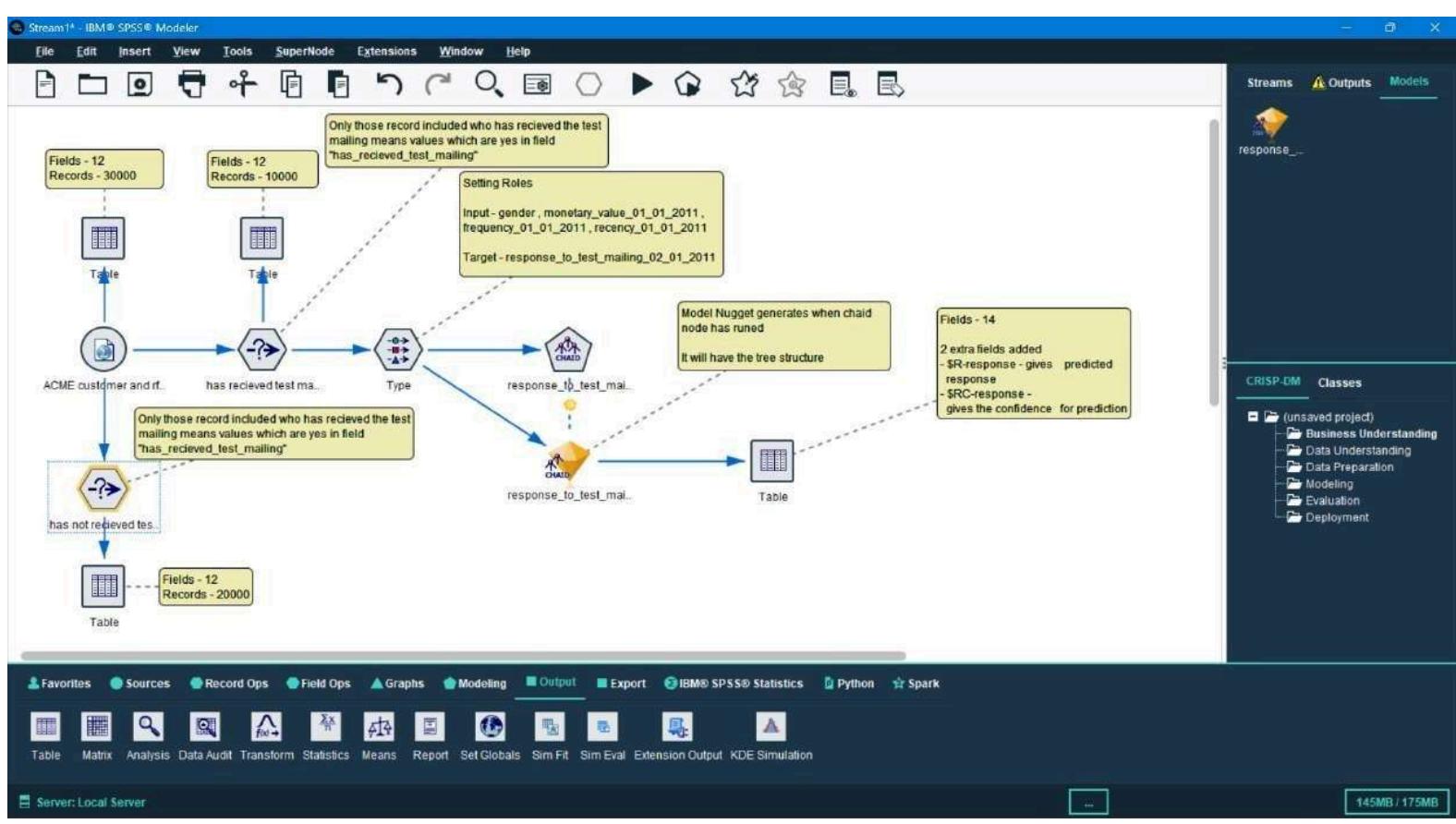
Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 130MB / 175MB

Table (14 fields, 10,000 records) #3

File Edit Generate

	_01_2011	has_received_test_mailing	response_to_test_mailing_02_01_2011	orderdate	number_of_days_between_mailing_and_orderdate	ordered_within_month	\$R-response_to_test_mailing_02_01_2011	SRC-response_to_test_mailing_02_01_2011	
1	yes	F	\$null\$		\$null\$	nap	F		0.953
2	yes	F	\$null\$		\$null\$	nap	F		0.993
3	yes	F	\$null\$		\$null\$	nap	F		0.953
4	yes	F	\$null\$		\$null\$	nap	F		0.953
5	yes	F	\$null\$		\$null\$	nap	F		0.993
6	yes	T	2011-02-04		3.000	yes	F		0.953
7	yes	F	\$null\$		\$null\$	nap	F		0.993
8	yes	F	\$null\$		\$null\$	nap	F		0.953
9	yes	F	\$null\$		\$null\$	nap	F		0.911
10	yes	T	2011-02-15		14.000	yes	T		0.625
11	yes	F	\$null\$		\$null\$	nap	F		0.911
12	yes	F	\$null\$		\$null\$	nap	F		0.953
13	yes	F	\$null\$		\$null\$	nap	F		0.998
14	yes	F	\$null\$		\$null\$	nap	F		0.911
15	yes	F	\$null\$		\$null\$	nap	F		0.953
16	yes	F	\$null\$		\$null\$	nap	F		0.993
17	yes	F	\$null\$		\$null\$	nap	F		0.953
18	yes	F	\$null\$		\$null\$	nap	F		0.911
19	res	F	\$null\$		\$null\$	nap	F		0.953
20	yes	F	\$null\$		\$null\$	nap	F		0.993
21	yes	F	\$null\$		\$null\$	nap	F		0.911
22	yes	T	2011-02-28		27.000	yes	F		0.953
23	yes	F	\$null\$		\$null\$	nap	F		0.911
24	yes	F	\$null\$		\$null\$	nap	F		0.993
25	yes	F	\$null\$		\$null\$	nap	F		0.998
26	yes	F	\$null\$		\$null\$	nap	F		1.000
27	yes	F	\$null\$		\$null\$	nap	F		0.998
28	yes	F	\$null\$		\$null\$	nap	F		0.999
29	yes	T	2011-02-17		16.000	yes	T		0.625
30	yes	F	\$null\$		\$null\$	nap	F		0.911
31	yes	F	\$null\$		\$null\$	nap	F		0.993
32	yes	F	\$null\$		\$null\$	nap	F		0.998
33	yes	T	2011-02-27		26.000	yes	F		0.953
34	yes	T	2011-02-22		21.000	yes	F		0.911
35	yes	F	\$null\$		\$null\$	nap	F		0.911
36	yes	F	\$null\$		\$null\$	nap	F		0.953
37	yes	F	\$null\$		\$null\$	nap	F		0.911
38	yes	F	\$null\$		\$null\$	nap	F		0.993
39	yes	F	\$null\$		\$null\$	nap	F		0.911
40	yes	F	\$null\$		\$null\$	nap	F		0.993
41	yes	T	2011-02-26		25.000	yes	T		0.625
42	yes	F	\$null\$		\$null\$	nap	F		0.953
43	res	F	\$null\$		\$null\$	nap	F		1.000



Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

The screenshot shows a data stream named "ACME customer and rf..." with two input nodes: "Fields - 12 Records - 30000" and "Fields - 12 Records - 10000". A third node, "Setting Roles", is connected between them. A "has not received test mailing" node is highlighted, with its properties dialog open. The dialog shows "Name: has not received test mailing", "Tooltip text:", and "Keywords:". It also displays annotations: "Only those record included who has received the test mailing means values which are yes in field 'has_received_test_mailing'", "Only those records included who has received the test mailing means values which are yes in field 'has_received_test_mailing'", "Setting Roles", "Input- gender, monetary_value_01_01_2011, frequency_01_01_2011, recency_01_01_2011", and "Fields - 14 2 extra fields added - \$R-response - gives predicted response - \$RC-response - gives the confidence for prediction". The stream continues with a "Table" node and ends at another "Table" node.

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 145MB / 175MB

Table (12 fields, 20,000 records) #2

File Edit Generate

	customer_id	gender	email_address	postal_code	monetary_value_01_01_2011	frequency_01_01_2011	recency_01_01_2011	has_received_test_mailing	response_to_test_mailing_02_01_2011	orderdate	number_of_days_between_mailing_ar
1	5724.000	male	name24503@wwmail.org	1530RE	1 low	3 high	1 low	no	\$null\$	\$null\$	\$null\$
2	5725.000	male	name18105@wwmail.es	1325CH	2 medium	2 medium	3 high	no	\$null\$	\$null\$	\$null\$
3	5726.000	female	name26276@wwmail.uk	1337YV	1 low	2 medium	2 medium	no	\$null\$	\$null\$	\$null\$
4	5727.000	female	name16698@wwmail.de	1503TD	2 medium	3 high	1 low	no	\$null\$	\$null\$	\$null\$
5	5729.000	male	name16699@wwmail.de	1688LE	1 low	2 medium	1 low	no	\$null\$	\$null\$	\$null\$
6	5730.000	male	name16701@wwmail.de	1424EE	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
7	5731.000	female	name1488@omejor.es	1123DZ	2 medium	2 medium	2 medium	no	\$null\$	\$null\$	\$null\$
8	5732.000	male	name18112@wwmail.es	1260NK	2 medium	1 low	2 medium	no	\$null\$	\$null\$	\$null\$
9	5734.000	male	name20733@wwmail.in	1913LG	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
10	5735.000	male	name10943@tnet.inc	1411MZ	1 low	3 high	1 low	no	\$null\$	\$null\$	\$null\$
11	5736.000	female	name956@hmail.uk	1457ZV	3 high	1 low	2 medium	no	\$null\$	\$null\$	\$null\$
12	5737.000	male	name16113@wwmail.es	1748JX	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
13	5739.000	male	name27890@zigzag.be	1794OK	1 low	3 high	2 medium	no	\$null\$	\$null\$	\$null\$
14	5740.000	male	name16715@wwmail.es	1365PV	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
15	5741.000	female	name13150@tnet.uk	1471AQ	1 low	3 high	2 medium	no	\$null\$	\$null\$	\$null\$
16	5742.000	female	name954@hmail.uk	1771PQ	1 low	2 medium	2 medium	no	\$null\$	\$null\$	\$null\$
17	5744.000	female	name6045@tnet.fr	1752EN	2 medium	2 medium	3 high	no	\$null\$	\$null\$	\$null\$
18	5745.000	female	name24499@wwmail.org	1738OB	2 medium	1 low	2 medium	no	\$null\$	\$null\$	\$null\$
19	5746.000	male	name1485@omejor.es	1259OY	3 high	1 low	2 medium	no	\$null\$	\$null\$	\$null\$
20	5747.000	female	name20726@wwmail.inc	1695OX	2 medium	1 low	2 medium	no	\$null\$	\$null\$	\$null\$
21	5749.000	female	name16725@wwmail.de	1111RW	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
22	5750.000	female	name24498@wwmail.org	1984NG	1 low	3 high	1 low	no	\$null\$	\$null\$	\$null\$
23	5751.000	male	name20723@wwmail.inc	1710KS	3 high	1 low	2 medium	no	\$null\$	\$null\$	\$null\$
24	5752.000	male	name24497@wwmail.org	1560CS	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
25	5754.000	male	name213@hmail.inc	1444MU	2 medium	3 high	2 medium	no	\$null\$	\$null\$	\$null\$
26	5755.000	female	name6074@tnet.fr	1013ZL	1 low	2 medium	1 low	no	\$null\$	\$null\$	\$null\$
27	5756.000	male	name26250@wwmail.uk	1982FK	1 low	3 high	2 medium	no	\$null\$	\$null\$	\$null\$
28	5757.000	female	name18120@wwmail.es	1512FH	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
29	5759.000	male	name28499@zigzag.cat	1876XQ	2 medium	2 medium	3 high	no	\$null\$	\$null\$	\$null\$
30	5760.000	male	name24495@wwmail.org	1392RD	1 low	2 medium	1 low	no	\$null\$	\$null\$	\$null\$
31	5761.000	female	name29808@zigzag.es	1057BH	1 low	2 medium	3 high	no	\$null\$	\$null\$	\$null\$
32	5762.000	male	name13148@tnet.uk	1015TR	2 medium	2 medium	3 high	no	\$null\$	\$null\$	\$null\$
33	5764.000	female	name4081@tnet.jp	1003UB	1 low	2 medium	3 high	no	\$null\$	\$null\$	\$null\$
34	5765.000	female	name27874@zigzag.be	1941JM	3 high	2 medium	2 medium	no	\$null\$	\$null\$	\$null\$
35	5766.000	male	name18125@wwmail.es	1682ME	2 medium	2 medium	2 medium	no	\$null\$	\$null\$	\$null\$
36	5767.000	female	name18132@wwmail.es	1611RI	1 low	3 high	1 low	no	\$null\$	\$null\$	\$null\$
37	5769.000	male	name10959@tnet.inc	1721ZF	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
38	5770.000	female	name18136@wwmail.es	1402SB	3 high	1 low	3 high	no	\$null\$	\$null\$	\$null\$
39	5771.000	male	name26243@wwmail.uk	1370LW	2 medium	1 low	3 high	no	\$null\$	\$null\$	\$null\$
40	5772.000	female	name6096@tnet.fr	1808FG	2 medium	2 medium	3 high	no	\$null\$	\$null\$	\$null\$
41	5774.000	female	name18143@wwmail.es	1037ZY	1 low	2 medium	1 low	no	\$null\$	\$null\$	\$null\$
42	5775.000	female	name20875@wwmail.inc	1465IB	2 medium	2 medium	1 low	no	\$null\$	\$null\$	\$null\$
43	5776.000	male	name18145@wwmail.es	1509KH	3 high	1 low	1 low	no	\$null\$	\$null\$	\$null\$

OK

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Setting Roles

Input - gender , monetary_value_01_01_2011 , frequency_01_01_2011 , recency_01_01_2011

Target - response_to_test_mailing_02_01_2011

ACME customer and rf... has received test mailing Type response_to_test_mailing

Only those record included who has received the test mailing means values which are "no" in field "has_received_test_mailing"

Model Nugget generates when chaid node has runned It will have the tree structure

Fields - 14 2 extra fields added - \$R-response - gives predicted response - \$RC-response - gives the confidence for prediction

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 159MB / 200MB

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Setting Roles

Input - gender , monetary_value_01_01_2011 , frequency_01_01_2011 , recency_01_01_2011

Target - response_to_test_mailing_02_01_2011

ACME customer and rf... has received test mailing Type response_to_test_mailing

Only those record included who has received the test mailing means values which are "no" in field "has_received_test_mailing"

predicted to buy

Mode: Include Discard

Condition:

Fields - 14 2 extra fields added - \$R-response - gives predicted response - \$RC-response - gives the confidence for prediction

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 159MB / 200MB

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

Streams Outputs Models

Fields - 12 Records - 30000 Fields - 12 Records - 10000 Only those record included who has received the test mailing means values which are "yes" in field "has_received_test_mailing".

Setting Roles
Input - gender , monetary_value_01_01_2011 , frequency_01_01_2011 , recency_01_01_2011
Target - response_to_test_mailing_02_01_2011

ACME customer and rf... has received test ma... Type
Only those record included who has received the test mailing means values which are "no" in field "has_received_test_mailing".

has not received tes... response_to_test_ma... predicted
Fields - 12 Records - 20000 Select only those record which have "1" value in field "SR-response".

predicted to buy

Settings Annotations
Name: Auto Custom predicted to buy
Tooltip text:
Keywords:
Created on 23 December, 2022 4:25 PM by dell. ID: jd3UGWTE9JH1F
This object has not been saved.

OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 159MB / 200MB

Table (14 fields, 254 records) #3

File Edit Generate

Table Annotations

	customer_id	gender	email_address	postal_code	monetary_value_01_01_2011	frequency_01_01_2011	recency_01_01_2011	has_received_test_mailing	response_to_test_mailing_02_01_2011	orderdate	number_of_days_between_mailing_ar
1	5835.000	male	name11005@thetlinc.	1573LT	3 high	2 medium	3 high	no	\$null\$	\$null\$	
2	5851.000	male	name16970@wwmail.de	1047IN	3 high	2 medium	3 high	no	\$null\$	\$null\$	
3	6031.000	female	name1440@lomejor.es	1874FC	3 high	2 medium	3 high	no	\$null\$	\$null\$	
4	6204.000	male	name1293@mail.uk	1607ZU	3 high	2 medium	3 high	no	\$null\$	\$null\$	
5	6245.000	female	name8288@thetlinc	1898WF	3 high	2 medium	3 high	no	\$null\$	\$null\$	
6	6340.000	female	name3721@mobca.cat	1868TM	3 high	2 medium	3 high	no	\$null\$	\$null\$	
7	6424.000	female	name1674@wwmail.de	1566LB	3 high	2 medium	3 high	no	\$null\$	\$null\$	
8	6451.000	male	name2299@wwmail.inc	1548VC	3 high	2 medium	3 high	no	\$null\$	\$null\$	
9	6460.000	male	name18645@wwmail.es	1627WH	3 high	2 medium	3 high	no	\$null\$	\$null\$	
10	6501.000	male	name1709@wwmail.de	1244PI	3 high	2 medium	3 high	no	\$null\$	\$null\$	
11	6516.000	female	name2252@wwmail.inc	1058JU	3 high	2 medium	3 high	no	\$null\$	\$null\$	
12	6525.000	female	name13046@thetlinc	1056TQ	3 high	2 medium	3 high	no	\$null\$	\$null\$	
13	6557.000	female	name13024@thetl.uk	1145YK	3 high	2 medium	3 high	no	\$null\$	\$null\$	
14	6559.000	male	name690@thetl.uk	1828GA	3 high	2 medium	3 high	no	\$null\$	\$null\$	
15	6766.000	female	name6252@thetl.fr	1158TL	3 high	2 medium	3 high	no	\$null\$	\$null\$	
16	6877.000	female	name8554@thetl.in	1677CH	3 high	2 medium	3 high	no	\$null\$	\$null\$	
17	6904.000	female	name15552@wwmail.de	1236UL	2 high	2 medium	3 high	no	\$null\$	\$null\$	
18	7162.000	female	name4301@thetl.jp	1857YS	3 high	2 medium	3 high	no	\$null\$	\$null\$	
19	7449.000	male	name2240@lomejor.es	1398WY	3 high	2 medium	3 high	no	\$null\$	\$null\$	
20	7510.000	male	name1935@wwmail.es	1097WR	3 high	2 medium	3 high	no	\$null\$	\$null\$	
21	7518.000	male	name2265@wwmail.inc	1280VW	3 high	2 medium	3 high	no	\$null\$	\$null\$	
22	7524.000	female	name1733@thetl.de	1407WB	3 high	2 medium	3 high	no	\$null\$	\$null\$	
23	7550.000	female	name1290@thetl.uk	1019CB	3 high	2 medium	3 high	no	\$null\$	\$null\$	
24	7555.000	female	name7288@thetl.fr	1169YI	3 high	2 medium	3 high	no	\$null\$	\$null\$	
25	7575.000	male	name2220@lomejor.es	1582JZ	3 high	2 medium	3 high	no	\$null\$	\$null\$	
26	7640.000	female	name19449@wwmail.com	1011UM	3 high	2 medium	3 high	no	\$null\$	\$null\$	
27	7694.000	male	name23168@wwmail.inc	1144EO	3 high	2 medium	3 high	no	\$null\$	\$null\$	
28	7810.000	male	name12862@thetl.uk	1105QV	3 high	2 medium	3 high	no	\$null\$	\$null\$	
29	7977.000	female	name16232@wwmail.de	1784XK	3 high	2 medium	3 high	no	\$null\$	\$null\$	
30	8002.000	female	name21184@wwmail.de	1442KG	3 high	2 medium	3 high	no	\$null\$	\$null\$	
31	8017.000	male	name2151@lomejor.es	1405DM	3 high	2 medium	3 high	no	\$null\$	\$null\$	
32	8149.000	male	name2133@lomejor.es	1591LQ	3 high	2 medium	3 high	no	\$null\$	\$null\$	
33	8160.000	male	name29738@zigzag.com	1505LE	3 high	2 medium	3 high	no	\$null\$	\$null\$	
34	8322.000	female	name19077@wwmail.es	1078ET	3 high	2 medium	3 high	no	\$null\$	\$null\$	
35	8356.000	male	name20918@zigzag.fr	1376WN	3 high	2 medium	3 high	no	\$null\$	\$null\$	
36	8378.000	male	name26202@wwmail.uk	1341XI	3 high	2 medium	3 high	no	\$null\$	\$null\$	
37	8429.000	male	name26310@wwmail.uk	1640PQ	3 high	2 medium	3 high	no	\$null\$	\$null\$	
38	8500.000	female	name28630@zigzag.ca	1831GE	3 high	2 medium	3 high	no	\$null\$	\$null\$	
39	8515.000	male	name1260@thetl.uk	1506Y	3 high	2 medium	3 high	no	\$null\$	\$null\$	
40	8557.000	male	name2101@wwmail.de	1098IP	3 high	2 medium	3 high	no	\$null\$	\$null\$	
41	8589.000	male	name29106@zigzag.jp	1105ZY	3 high	2 medium	3 high	no	\$null\$	\$null\$	
42	8789.000	female	name26982@wwmail.uk	1155CN	3 high	2 medium	3 high	no	\$null\$	\$null\$	
43	8799.000	female	name1935d@wwmail.uk	1650XI	3 high	2 medium	3 high	no	\$null\$	\$null\$	

OK

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Fields - 12 Records - 30000 Fields - 12 Records - 10000 Only those record included who has received the test mailing means values which are "yes" in field "has_received_test_mailing".

Setting Roles
Input - gender, monetary_value_01_01_2011, frequency_01_01_2011, recency_01_01_2011
Target - response_to_test_mailing_02_01_2011

ACME customer and rf... has received test ma... Type response_to_test_ma... CHAID response_to_test_ma... Fields - 14
Only those record included who has received the test mailing means values which are "no" in field "has_received_test_mailing".
Model Nugget generates when chaid node has runned
It will have the tree structure.
2 extra fields added
- \$R-response - gives predicted response
- \$RC-response - gives the confidence for prediction.

has not received tes... response_to_test_ma... response_to_test_ma... predicted to buy Number of customer to the product - 254
Fields - 12 Records - 20000
Table
Select only those record which have "T" value in field "\$R-response".
All fields filtered out except customer_id, \$R-response, \$RC-response.

Favorites Sources Record Ops Field Ops Modeling Output IBM® SPSS® Statistics Python Spark

Auto Data Prep Type Filter Derive Filter Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproj. Time Intervals

Server: Local Server 164MB / 200MB

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Fields - 12 Records - 30000 Fields - 12 Records - 10000 Only those record included who has received the test mailing means values which are "yes" in field "has_received_test_mailing".

Setting Roles
Input - gender, monetary_value_01_01_2011, frequency_01_01_2011, recency_01_01_2011
Target - response_to_test_mailing_02_01_2011

ACME customer and rf... has received test ma... Type response_to_test_ma... CHAID response_to_test_ma... Fields - 14
Only those record included who has received the test mailing means values which are "no" in field "has_received_test_mailing".
Model Nugget generates when chaid node has runned
It will have the tree structure.
2 extra fields added
- \$R-response - gives predicted response
- \$RC-response - gives the confidence for prediction.

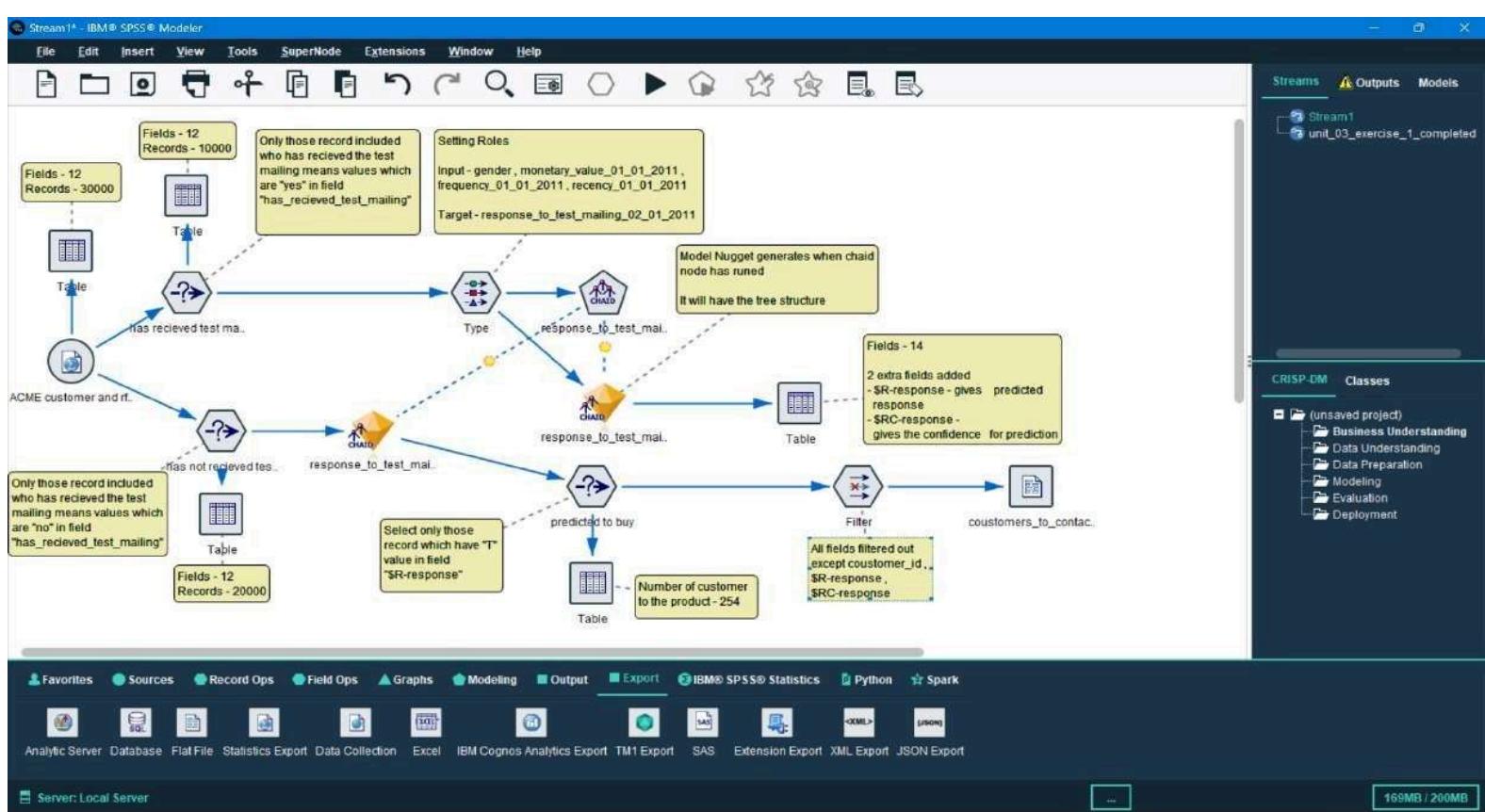
has not received tes... response_to_test_ma... response_to_test_ma... predicted to buy Number of customer to the product - 254
Fields - 12 Records - 20000
Table
Select only those record which have "T" value in field "\$R-response".
All fields filtered out except customer_id, \$R-response, \$RC-response.

Filter dialog box: Filter Annotations
Fields: 14 in, 11 filtered, 2 renamed, 3 out
Field Filter Field
monetary_value_01_01_2011 → monetary_value_01_01_2011
frequency_01_01_2011 → frequency_01_01_2011
recency_01_01_2011 → recency_01_01_2011
has_received_test_mailing → has_received_test_mailing
response_to_test_mailing_0... → response_to_test_mailing_0...
orderedate → orderedate
number_of_days_between_... → number_of_days_between_...
ordered_within_month → ordered_within_month
\$R-response_to_test_mailing... → predicted_category
\$RC-response_to_test_mailing... → confidence_for_prediction
View current fields (radio button selected) View unused field settings
OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops Modeling Output IBM® SPSS® Statistics Python Spark

Auto Data Prep Type Filter Derive Filter Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproj. Time Intervals

Server: Local Server 164MB / 200MB



customers_to_contact - Notepad

customer_id	predicted_category	confidence_for_prediction
5835.000	"T"	0.625
5851.000	"T"	0.625
6031.000	"T"	0.625
6204.000	"T"	0.625
6245.000	"T"	0.625
6340.000	"T"	0.625
6424.000	"T"	0.625
6451.000	"T"	0.625
6460.000	"T"	0.625
6501.000	"T"	0.625
6516.000	"T"	0.625
6525.000	"T"	0.625
6657.000	"T"	0.625
6659.000	"T"	0.625
6766.000	"T"	0.625
6877.000	"T"	0.625
6904.000	"T"	0.625
7162.000	"T"	0.625
7449.000	"T"	0.625
7510.000	"T"	0.625
7516.000	"T"	0.625
7524.000	"T"	0.625
7550.000	"T"	0.625
7555.000	"T"	0.625
7576.000	"T"	0.625
7640.000	"T"	0.625
7694.000	"T"	0.625
7810.000	"T"	0.625
7977.000	"T"	0.625
8002.000	"T"	0.625
8017.000	"T"	0.625
8149.000	"T"	0.625
8160.000	"T"	0.625
8322.000	"T"	0.625
8356.000	"T"	0.625
8376.000	"T"	0.625
8429.000	"T"	0.625
8500.000	"T"	0.625
8515.000	"T"	0.625
8557.000	"T"	0.625
8589.000	"T"	0.625

Ln 1, Col 1 100% Windows (CRLF) UTF-8

PROJECT-2

Collecting Initial Data

Stream2* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream2 Exercise-1

COUSTOMERS
RECORDS- 30000
FIELDS - 1
UNIT OF ANALYSIS -
CUSTOMER_ID

SETTING
MEASUREMENT
LEVEL

ACME customers.xlsx Type Table 1 RECORD PER CUSTOMER

CRISP-DM Classes

(unsaved project)
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS® Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 153MB / 214MB

Stream2* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream2 Exercise-1

COUSTOMERS
RECORDS- 30000
FIELDS - 1
UNIT OF ANALYSIS -
CUSTOMER_ID

SETTING
MEASUREMENT
LEVEL

ACME customers.xlsx Type

Type dialog box:

Field	Measurement	Values	Missing	Check	Role
CUSTOMER_ID	Typeless			None	<input type="checkbox"/> None
GENDER	Nominal	F,M,f,m		None	<input checked="" type="checkbox"/> Input
CREDITLIMIT	Continuous	-999999.0...		None	<input checked="" type="checkbox"/> Input
ZODIAC	Nominal	1.0,2.0,3.0...		None	<input checked="" type="checkbox"/> Input
E-MAIL ADD...	Typeless			None	<input type="checkbox"/> None
ZIP	Typeless			None	<input type="checkbox"/> None

CRISP-DM Classes

(unsaved project)
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS® Statistics Python Spark

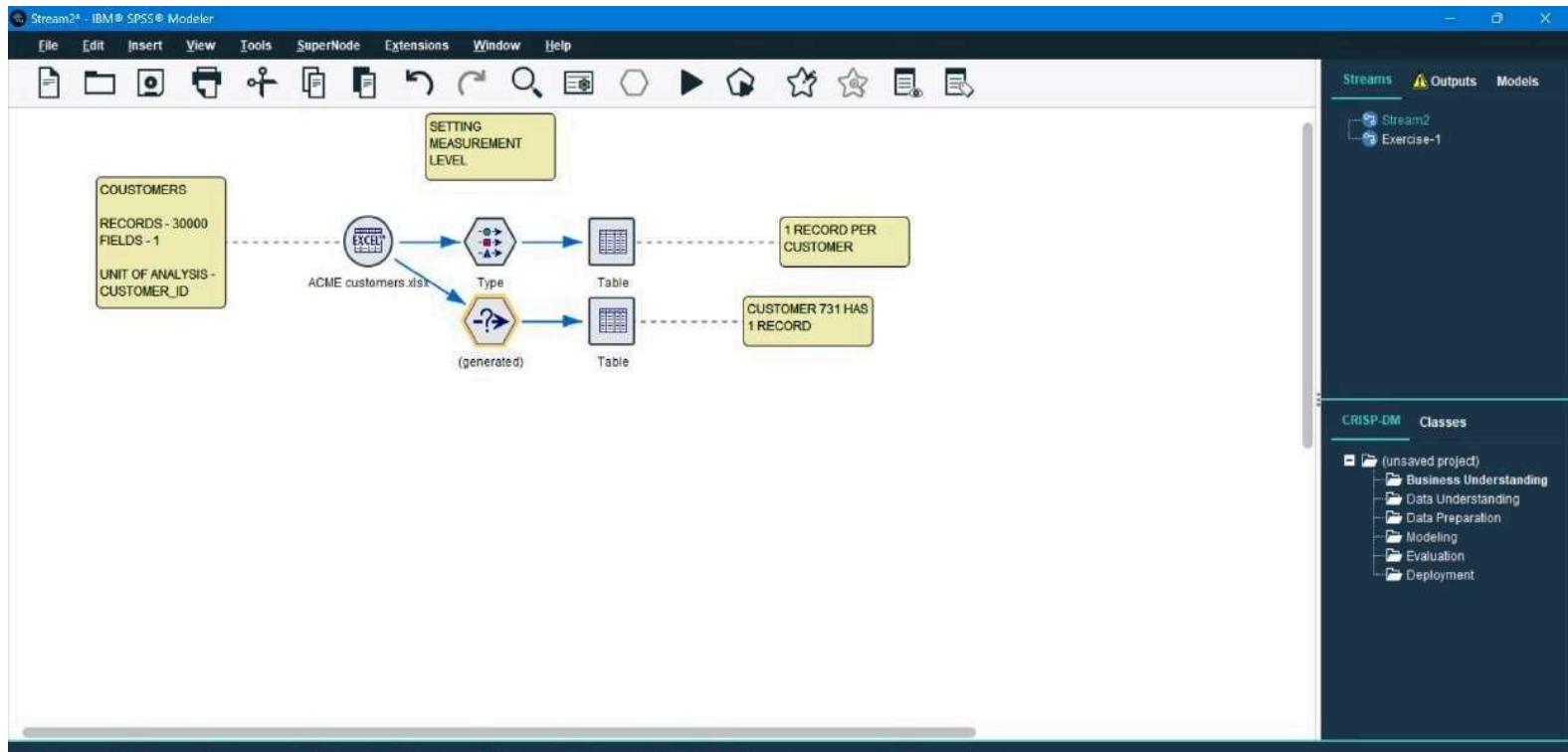
Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 154MB / 214MB

Table (6 fields, 30,000 records)

Annotations

	CUSTOMER_ID	GENDER	CREDITLIMIT	ZODIAC	E-MAIL ADDRESS	ZIP
1	723.000 F	9026.000	1.000	name7502@het.fr	1618B0	
2	724.000 M	5223.000	3.000	name25485@wwwmail.org	1132DG	
3	725.000 F	668.000	8.000	name15643@wwwmail.de	1803YT	
4	726.000 M	8302.000	6.000	name28335@zuzag.be	1205WR	
5	727.000 F	997.000	8.000	name5354@het.jp	1711ON	
6	728.000 M	1955.000	3.000	name20637@wwwmail.es	1055FG	
7	729.000 F	6898.000	5.000	name20636@wwwmail.es	1254MR	
8	730.000 F	6488.000	11.000	name10414@het.net.inc	1723DG	
9	731.000 F	5006.000	2.000	name23372@wwwmail.in	1713AQ	
10	732.000 F	929.000	5.000	name20635@wwwmail.es	1284EC	
11	733.000 F	3303.000	8.000	name5356@het.jp	1648BT	
12	734.000 F	6934.000	3.000	name17582@wwwmail.de	1285VX	
13	735.000 F	9808.000	7.000	name388@het.fr	1282NB	
14	736.000 M	4723.000	6.000	name10409@het.net.inc	1799IT	
15	737.000 M	672.000	7.000	name13849@het.uk	1802UO	
16	738.000 F	7801.000	8.000	name25473@wwwmail.org	1971NK	
17	739.000 M	5984.000	5.000	name13848@het.uk	1361RL	
18	740.000 F	2621.000	9.000	name23356@wwwmail.inc	1164VN	
19	741.000 M	6062.000	8.000	name3188@molbe.ca	1767YN	
20	742.000 M	9797.000	4.000	name1606@lomejor.es	1681HP	
21	743.000 F	2687.000	6.000	name29056@zuzag.cat	1287PI	
22	744.000 M	128.000	9.000	name28323@zuzag.be	1023AR	
23	745.000 F	182.000	11.000	name20630@wwwmail.es	1737ZI	
24	746.000 M	2301.000	9.000	name13847@het.uk	1625TZ	
25	747.000 F	2704.000	8.000	name10401@het.net.inc	1684YE	
26	748.000 M	4033.000	7.000	name13846@het.uk	1508DU	
27	749.000 M	1818.000	9.000	name17579@wwwmail.de	1143VC	
28	750.000 F	9044.000	5.000	name26994@wwwmail.uk	1229IK	
29	751.000 F	1975.000	12.000	name5366@het.fr	1567CX	
30	752.000 F	3619.000	9.000	name1609@lomejor.es	1315FI	
31	753.000 M	3517.000	8.000	name29727@zuzag.com	1753OI	
32	754.000 M	701.000	8.000	name29656@zuzag.es	1229DC	
33	755.000 M	1878.000	3.000	name17578@wwwmail.de	1036LJ	
34	756.000 M	4023.000	10.000	name29046@zuzag.cat	1002VW	
35	757.000 M	4946.000	10.000	name10380@het.net.inc	1979KZ	
36	758.000 F	9946.000	8.000	name5354@het.fr	1784IE	
37	759.000 F	1522.000	6.000	name3172@molbe.cat	1183FC	
38	760.000 F	7458.000	2.000	name43@mail.de	1492WX	
39	761.000 F	1340.000	10.000	name5357@het.jp	1569VY	
40	762.000 F	7136.000	3.000	name255@mail.inc	1558OG	
41	763.000 F	1723.000	4.000	name25457@wwwmail.org	1588SB	
42	764.000 F	3799.000	7.000	name1616@lomejor.es	1972BQ	
43	765.000 M	1326.000	2.000	name5359@het.jp	1109VD	
44	766.000 F	8432.000	2.000	name17576@wwwmail.de	1034PF	



Stream2 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

COUSTOMERS
RECORDS - 30000
FIELDS - 1
UNIT OF ANALYSIS - CUSTOMER_ID

SETTING MEASUREMENT LEVEL

ACME customers.xlsx

Type Table

1 RECORD PER CUSTOMER

(generated)

Preview

Settings Annotations

Mode: Include Discard

Condition: CUSTOMER_ID = 731.0

OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 176MB / 214MB

Stream2 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

COUSTOMERS
RECORDS - 30000
FIELDS - 1
UNIT OF ANALYSIS - CUSTOMER_ID

SETTING MEASUREMENT LEVEL

ACME customers.xlsx

Type Table

1 RECORD PER CUSTOMER

(generated) Table

CUSTOMER 731 HAS 1 RECORD

Table (6 fields, 1 records)

File Edit Generate

Annotations

CUSTOMER_ID	GENDER	CREDITLIMIT	ZODIAC	E-MAIL ADDRESS	ZIP
1	731.000.F	5006.000.	2.000	name23372@wwwmail.in	1713AQ

OK

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 175MB / 214MB

Stream2* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS® Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 170MB / 246MB

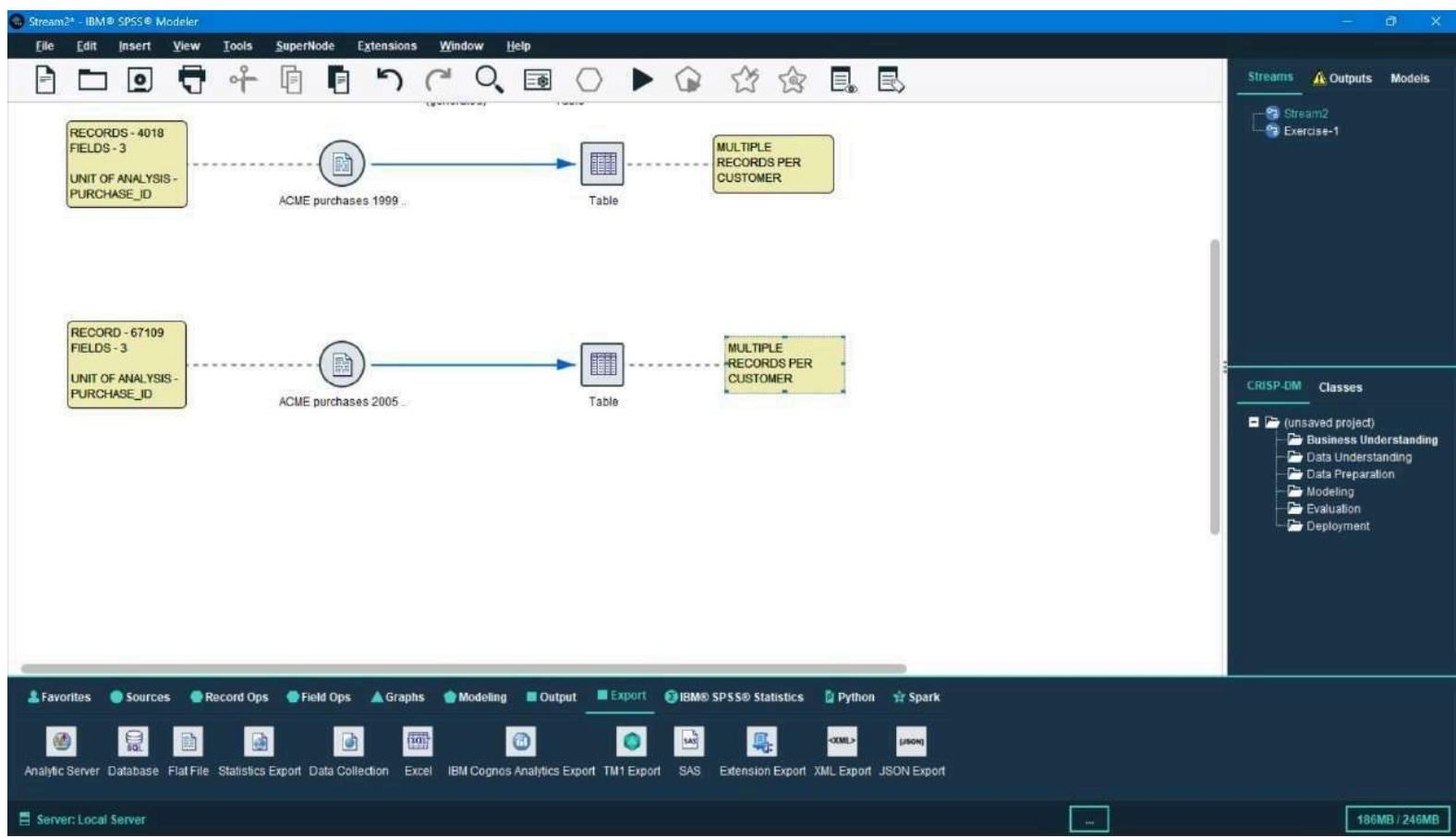
Table (3 fields, 4,018 records)

File Edit Generate

Table Annotations

	CUSTOMER_ID	PURCHASE_ID	ORDERDATE
1	724	80371999-07-11	
2	727	93241999-02-21	
3	728	103852004-05-22	
4	730	69662002-11-29	
5	731	86202000-07-21	
6	731	63362001-11-09	
7	731	75832003-07-17	
8	732	106482003-04-22	
9	732	83302003-05-10	
10	733	106182001-05-01	
11	733	58382003-12-01	
12	736	82962000-11-13	
13	736	63572003-10-22	
14	737	74532003-03-09	
15	738	60252002-11-13	
16	739	62532003-01-28	
17	740	81921999-12-07	
18	740	69022001-02-05	
19	741	95541999-07-15	
20	741	81432001-10-04	
21	742	59872000-01-22	
22	743	102452001-01-22	
23	743	71722003-09-16	
24	744	183862004-08-09	
25	745	85322001-08-24	
26	745	87652001-10-20	
27	748	72722000-01-20	
28	748	65402003-08-27	
29	751	101371999-01-09	
30	751	72091999-10-24	
31	752	69812000-04-10	
32	752	95652002-12-13	
33	753	82272002-01-18	
34	755	183872004-05-14	
35	756	87742004-04-16	
36	756	97212004-09-15	
37	756	183882004-11-15	
38	757	106832000-05-01	
39	759	82522001-08-15	
40	759	93302003-03-04	
41	760	105822001-11-19	
42	761	64921999-12-06	
43	761	68672002-04-18	
44	761	50052002-11-24	

OK



Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 186MB / 246MB

Table (3 fields, 67,109 records)

	CUSTOMER_ID	PURCHASE_ID	ORDERD...
1	723	707867026	2008-03-01
2	723	707867024	2009-07-02
3	725	708270082	2005-09-14
4	725	708270077	2005-09-24
5	725	708270079	2006-11-08
6	725	708270086	2007-02-27
7	726	708477065	2005-01-14
8	726	708477054	2005-08-27
9	726	708477050	2007-12-26
10	728	708477059	2008-04-11
11	729	708477045	2010-01-29
12	726	708477047	2010-06-14
13	729	708569840	2005-07-13
14	734	708595177	2008-09-28
15	734	708595174	2010-03-12
16	734	708595172	2010-07-28
17	735	708637782	2005-01-12
18	735	708637784	2005-11-17
19	735	708637787	2009-05-26
20	748	708647052	2005-04-29
21	748	708647050	2008-01-28
22	747	708637955	2008-08-15
23	749	708974921	2007-11-20
24	749	708974919	2009-02-08
25	750	709124899	2005-07-09
26	750	709124896	2009-02-03
27	750	709124894	2009-02-23
28	754	709160989	2007-04-12
29	754	709160991	2010-04-29
30	758	709217404	2005-01-27
31	758	709217402	2008-01-23
32	758	709217407	2010-10-30
33	763	709241931	2011-06-30
34	767	709402000	2009-05-04
35	768	709540470	2008-10-30
36	769	709703126	2005-04-12
37	769	709703124	2010-08-22
38	771	709832508	2005-03-14
39	771	709832503	2006-08-02
40	771	709832496	2008-04-28
41	771	709832514	2009-04-27
42	771	709832499	2009-11-07
43	771	709832494	2010-11-15
44	772	709832490	2007-12-28

OK

Stream2 - IBM SPSS Modeler

File Edit Insert View Tools SuperNode Window Help

Streams Outputs Models

Exercise-1

RECORDS - 4018
FIELDS - 3
UNIT OF ANALYSIS - PURCHASE_ID

PURCHASES

ACME purchases 1999...

MULTIPLE RECORDS PER CUSTOMER

Table

(generated)

CUSTOMER 731 HAS 3 RECORDS

RECORD - 67109
FIELDS - 3
UNIT OF ANALYSIS - PURCHASE_ID

ACME purchases 2005

MULTIPLE RECORDS PER CUSTOMER

Table

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 187MB / 247MB

```
graph LR; PURCHASES[PURCHASES] -- "ACME purchases 1999..." --> T1[Table]; PURCHASES -- "ACME purchases 2005" --> T2[Table]; T1 -- "(generated)" --> T2; T1 --> MRPC1[MULTIPLE RECORDS PER CUSTOMER]; T2 --> MRPC2[MULTIPLE RECORDS PER CUSTOMER]; MRPC1 --> CRISPDM[CRISP-DM Classes]; MRPC2 --> CRISPDM;
```

Table (3 fields, 3 records) #1

File Edit Generate Help

Annotations

	CUSTOMER_ID	PURCHASE_ID	ORDERDATE
1	731	8620	2000-07-21
2	731	6336	2001-11-09
3	731	7583	2003-07-17

MULTIPLE RECORDS PER CUSTOMER

Table

(generated)

CUSTOMER 731 HAS 3 RECORDS

OK

RECORD - 67109
FIELDS - 3
UNIT OF ANALYSIS - PURCHASE_ID

ACME purchases 2005

MULTIPLE RECORDS PER CUSTOMER

Table

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 188MB / 247MB

```
graph LR; T1[Table] -- "ACME purchases 1999..." --> T2[Table]; T1 -- "ACME purchases 2005" --> T2; T1 --> MRPC1[MULTIPLE RECORDS PER CUSTOMER]; T2 --> MRPC2[MULTIPLE RECORDS PER CUSTOMER]; MRPC1 --> CRISPDM[CRISP-DM Classes]; MRPC2 --> CRISPDM;
```

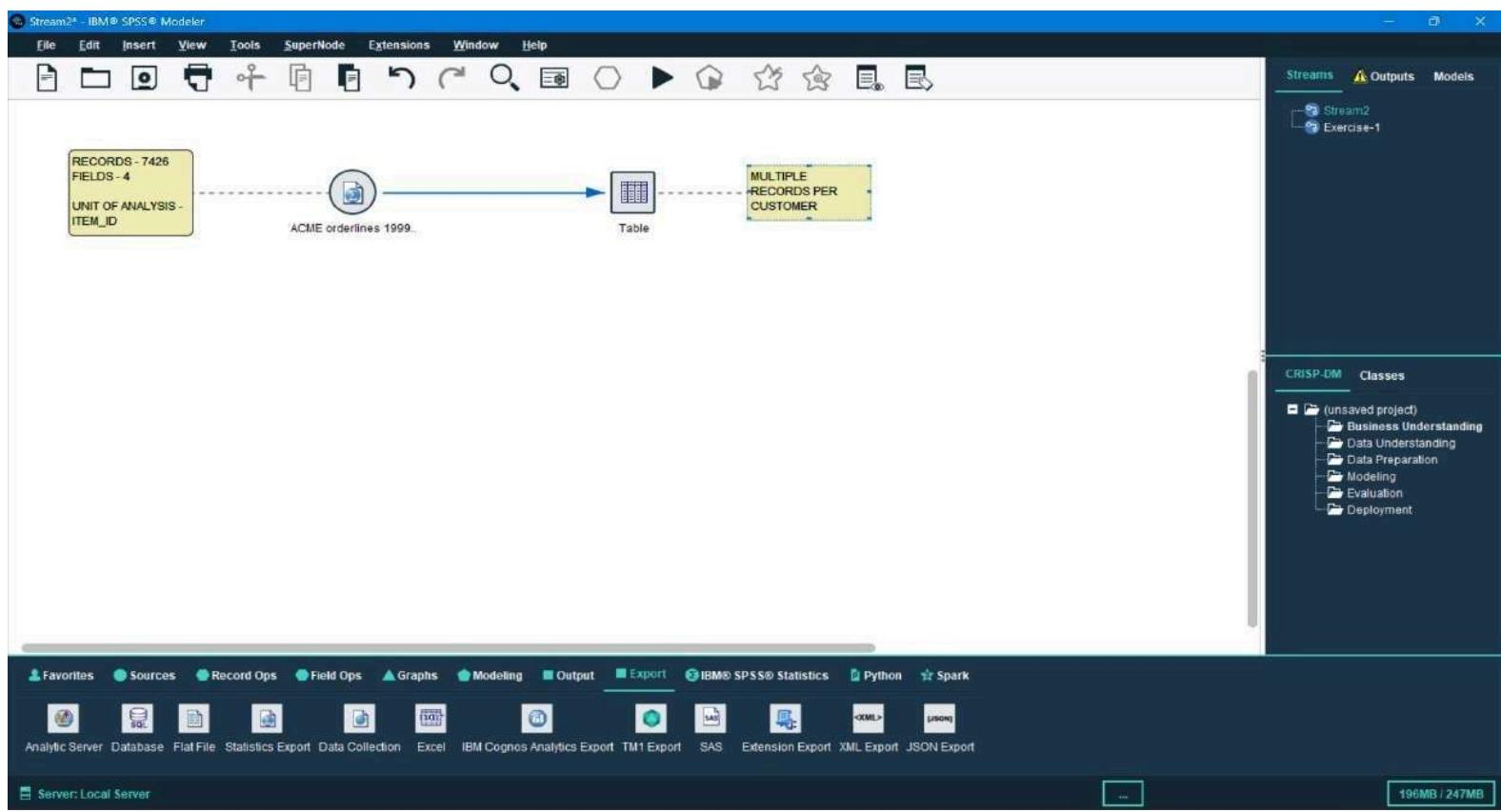


Table (4 fields, 7,426 records)

File Edit Generate ? X

Table Annotations

	PURCHASE_ID	ITEM_ID	PRODUCT_ID	PRICE
1	5723.000	16243...	10726.000	197.040
2	5723.000	16643...	10736.000	143.280
3	5723.000	15183...	10746.000	129.840
4	5723.000	12442...	10748.000	478.320
5	5724.000	15244...	10732.000	114.840
6	5726.000	17640...	10725.000	217.770
7	5727.000	14843...	10723.000	103.480
8	5727.000	11494...	10733.000	164.100
9	5727.000	10942...	10742.000	198.550
10	5729.000	17131...	10723.000	243.480
11	5729.000	18058...	10730.000	279.050
12	5729.000	17320...	10732.000	462.880
13	5730.000	17384...	10738.000	672.390
14	5730.000	14449...	10752.000	262.960
15	5731.000	13120...	10731.000	43.940
16	5731.000	18201...	10736.000	702.900
17	5732.000	14202...	10731.000	45.620
18	5732.000	15940...	10736.000	311.040
19	5732.000	14529...	10747.000	260.980
20	5732.000	13400...	10748.000	322.920
21	5734.000	16808...	10730.000	460.000
22	5736.000	14193...	10724.000	296.950
23	5736.000	13599...	10747.000	308.460
24	5737.000	18262...	10723.000	413.350
25	5737.000	15915...	10725.000	486.900
26	5737.000	18351...	10731.000	299.320
27	5738.000	15948...	10734.000	298.370
28	5739.000	14598...	10742.000	30.450
29	5740.000	14587...	10737.000	201.880
30	5741.000	15278...	10741.000	319.200
31	5743.000	13011...	10747.000	39.570
32	5744.000	14427...	10744.000	75.520
33	5745.000	13764...	10747.000	60.560
34	5746.000	13768...	10729.000	389.000
35	5746.000	16540...	10738.000	112.280
36	5747.000	16269...	10737.000	298.700
37	5747.000	16460...	10752.000	228.080
38	5748.000	16546...	10736.000	274.400
39	5749.000	15610...	10739.000	803.840
40	5748.000	15387...	10745.000	165.060
41	5748.000	17958...	10751.000	35.970
42	5749.000	15579...	10740.000	288.630
43	5751.000	14353...	10724.000	542.400
44	5751.000	17425...	10749.000	202.440

OK

Stream2* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream2 Exercise-1

RECORDS - 7426
FIELDS - 4
UNIT OF ANALYSIS - ITEM_ID

ACME orderlines 1999

MULTIPLE RECORDS PER CUSTOMER

PURCHASE_ID 6339 OF CUSTOMER 731 HAS 3 RECORDS

Table (generated)

Table

CRISP-DM Classes

(unsaved project)
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 214MB / 247MB

```
graph LR; Source[RECORDS - 7426  
FIELDS - 4  
UNIT OF ANALYSIS - ITEM_ID] --> ACME[ACME orderlines 1999]; ACME --> Generated((generated)); Generated --> Table1[Table]; Generated --> Table2[Table]; Table1 --> MultiRecords[MULTIPLE RECORDS PER CUSTOMER]; Table2 --> Purchase[PURCHASE_ID 6339  
OF CUSTOMER 731  
HAS 3 RECORDS]
```

Stream2* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream2 Exercise-1

RECORDS - 7426
FIELDS - 4
UNIT OF ANALYSIS - ITEM_ID

ACME orderlines 1999

MULTIPLE RECORDS PER CUSTOMER

PURCHASE_ID = 6336.0

Table (generated)

Settings Annotations

Mode: Include Discard

Condition:

OK Cancel Apply Reset

CRISP-DM Classes

(unsaved project)
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 214MB / 247MB

```
graph LR; Source[RECORDS - 7426  
FIELDS - 4  
UNIT OF ANALYSIS - ITEM_ID] --> ACME[ACME orderlines 1999]; ACME --> Generated((generated)); Generated --> Table1[Table]; Generated --> Table2[Table]; Table1 --> MultiRecords[MULTIPLE RECORDS PER CUSTOMER]; Table2 --> Purchase[PURCHASE_ID = 6336.0]
```

Stream2 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream2 Exercise-1

RECORDS - 7426 FIELDS - 4 UNIT OF ANALYSIS - ITEM_ID

ACME orderlines 1999.

(generated)

MULTIPLE RECORDS PER CUSTOMER

PURCHASE_ID 6336 OF CUSTOMER 731 HAS 3 RECORDS

Table (4 fields, 3 records)

Table Annotations

	PURCHASE_ID	ITEM_ID	PRODUCT_ID	PRICE
1	6336.000	12976...	10724.000	951.9...
2	6336.000	17485...	10726.000	12.480
3	6336.000	15205...	10727.000	587.7...

OK

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 215MB / 247MB

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Stream2 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream2 Exercise-1

RECORDS - 7426 FIELDS - 4 UNIT OF ANALYSIS - ITEM_ID

ACME orderlines 1999.

(generated)

MULTIPLE RECORDS PER CUSTOMER

PURCHASE_ID 6336 OF CUSTOMER 731 HAS 3 RECORDS

ORDERLINES

RECORDS - 162308 FIELDS - 4 UNIT OF ANALYSIS - ITEM_ID

ACME orderlines 2005.

Table (4 fields, 3 records)

Table Annotations

	PURCHASE_ID	ITEM_ID	PRODUCT_ID	PRICE
1	6336.000	12976...	10724.000	951.9...
2	6336.000	17485...	10726.000	12.480
3	6336.000	15205...	10727.000	587.7...

OK

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 172MB / 249MB

CRISP-DM Classes

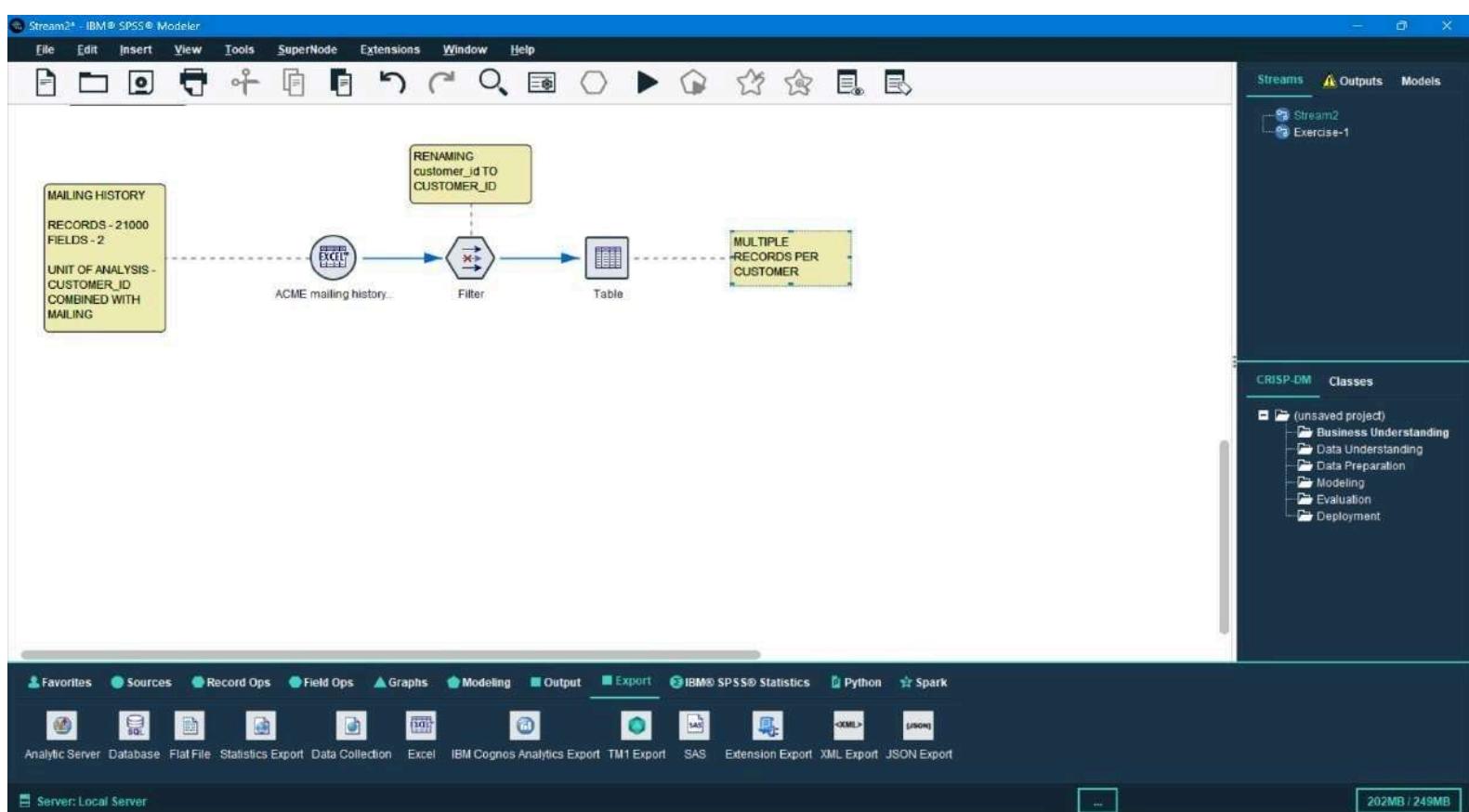
- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Table (4-Fields, 162,306 records)

File Edit Generate Search

Annotations

	PURCHASE_ID	ITEM_ID	PRODUCT_ID	PRICE
1	707867024.0...	4727541321.0...	10735.000	69.521
2	707867024.0...	5933029444.0...	10741.000	152.308
3	707867026.0...	3349835079.0...	10734.000	324.903
4	708270077.0...	5145284958.0...	10734.000	73.191
5	708270077.0...	4679332536.0...	10739.000	452.667
6	708270079.0...	4450722394.0...	10724.000	256.858
7	708270082.0...	5977515801.0...	10723.000	271.570
8	708270082.0...	2038221846.0...	10742.000	283.599
9	708270082.0...	7024719200.0...	10745.000	244.274
10	708270082.0...	2561070747.0...	10746.000	90.724
11	708270086.0...	5985143962.0...	10724.000	107.695
12	708270086.0...	4438932721.0...	10750.000	35.437
13	708477045.0...	2017033366.0...	10745.000	33.212
14	708477045.0...	3863474485.0...	10749.000	80.500
15	708477045.0...	4381966403.0...	18383.000	50.000
16	708477047.0...	1939110097.0...	18383.000	75.000
17	708477050.0...	5879159711.0...	10745.000	146.493
18	708477050.0...	2399497092.0...	10752.000	235.673
19	708477054.0...	3701455016.0...	10729.000	156.625
20	708477054.0...	3720961749.0...	10739.000	333.576
21	708477054.0...	5544363924.0...	10741.000	105.901
22	708477054.0...	4625316578.0...	10746.000	349.805
23	708477059.0...	4266549958.0...	10725.000	141.723
24	708477059.0...	4046856721.0...	10733.000	88.877
25	708477059.0...	4068117415.0...	10736.000	345.452
26	708477065.0...	6372163973.0...	10729.000	73.134
27	708477065.0...	3535168910.0...	10733.000	20.263
28	708477065.0...	2204575459.0...	10740.000	146.247
29	708477065.0...	4943392642.0...	10750.000	106.690
30	708569840.0...	3623648517.0...	10723.000	34.724
31	708595172.0...	6745273204.0...	10723.000	235.533
32	708595172.0...	3574783507.0...	10745.000	195.113
33	708595172.0...	2830792497.0...	18383.000	100.000
34	708595174.0...	3588525877.0...	10736.000	75.281
35	708595177.0...	6620956378.0...	10737.000	192.626
36	708595177.0...	5020531755.0...	10739.000	23.204
37	708595177.0...	5958464602.0...	10747.000	115.861
38	708595177.0...	6118078192.0...	10751.000	98.409
39	7086337782.0...	3074727933.0...	10741.000	14.349
40	7086337782.0...	4659916353.0...	10743.000	70.352
41	7086337784.0...	2285587609.0...	10747.000	127.482
42	7086337784.0...	5936234157.0...	10749.000	20.292
43	7086337787.0...	4204197122.0...	10748.000	138.904
44	7086337787.0...	5133825073.0...	10752.000	85.508



Stream2 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

MAILING HISTORY
RECORDS - 21000
FIELDS - 2
UNIT OF ANALYSIS -
CUSTOMER_ID
COMBINED WITH
MAILING

RENAMING
customer_id TO
CUSTOMER_ID

ACME mailing history...

Filter

Preview

Annotations

Fields: 2 in, 0 filtered, 0 renamed, 2 out

Field	Filter	Field
CUSTOMER_ID	→	CUSTOMER_ID
mailing	→	mailing

View current fields View unused field settings

OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 203MB / 249MB

Table (2 fields, 21,000 records)

File Edit Generate

Annotations

	CUSTOMER_ID	mailing
1	723.000	XL original orange baseball cap
2	723.000	XS original red soccer shirt
3	724.000	XL original orange baseball cap
4	725.000	XL original orange baseball cap
5	726.000	XL original orange baseball cap
6	727.000	Standard tennisracket
7	727.000	XL original orange baseball cap
8	727.000	XS original red soccer shirt
9	728.000	XL original orange baseball cap
10	729.000	XL original orange baseball cap
11	730.000	Standard tennisracket
12	730.000	XL original orange baseball cap
13	731.000	XL original orange baseball cap
14	731.000	XS original red soccer shirt
15	732.000	XL original orange baseball cap
16	732.000	XS original red soccer shirt
17	733.000	XL original orange baseball cap
18	733.000	XS original red soccer shirt
19	734.000	XL original orange baseball cap
20	734.000	XS original red soccer shirt
21	735.000	XL original orange baseball cap
22	736.000	XL original orange baseball cap
23	736.000	XS original red soccer shirt
24	737.000	Standard tennisracket
25	737.000	XL original orange baseball cap
26	738.000	XL original orange baseball cap
27	739.000	XL original orange baseball cap
28	740.000	Standard tennisracket
29	740.000	XL original orange baseball cap
30	741.000	XL original orange baseball cap
31	742.000	XL original orange baseball cap
32	742.000	XS original red soccer shirt
33	743.000	XL original orange baseball cap
34	743.000	XS original red soccer shirt
35	744.000	XL original orange baseball cap
36	745.000	XL original orange baseball cap
37	746.000	Standard tennisracket
38	746.000	XL original orange baseball cap
39	747.000	Standard tennisracket
40	747.000	XL original orange baseball cap
41	748.000	XL original orange baseball cap
42	749.000	XL original orange baseball cap
43	750.000	XL original orange baseball cap
44	751.000	XL original orange baseball cap

OK

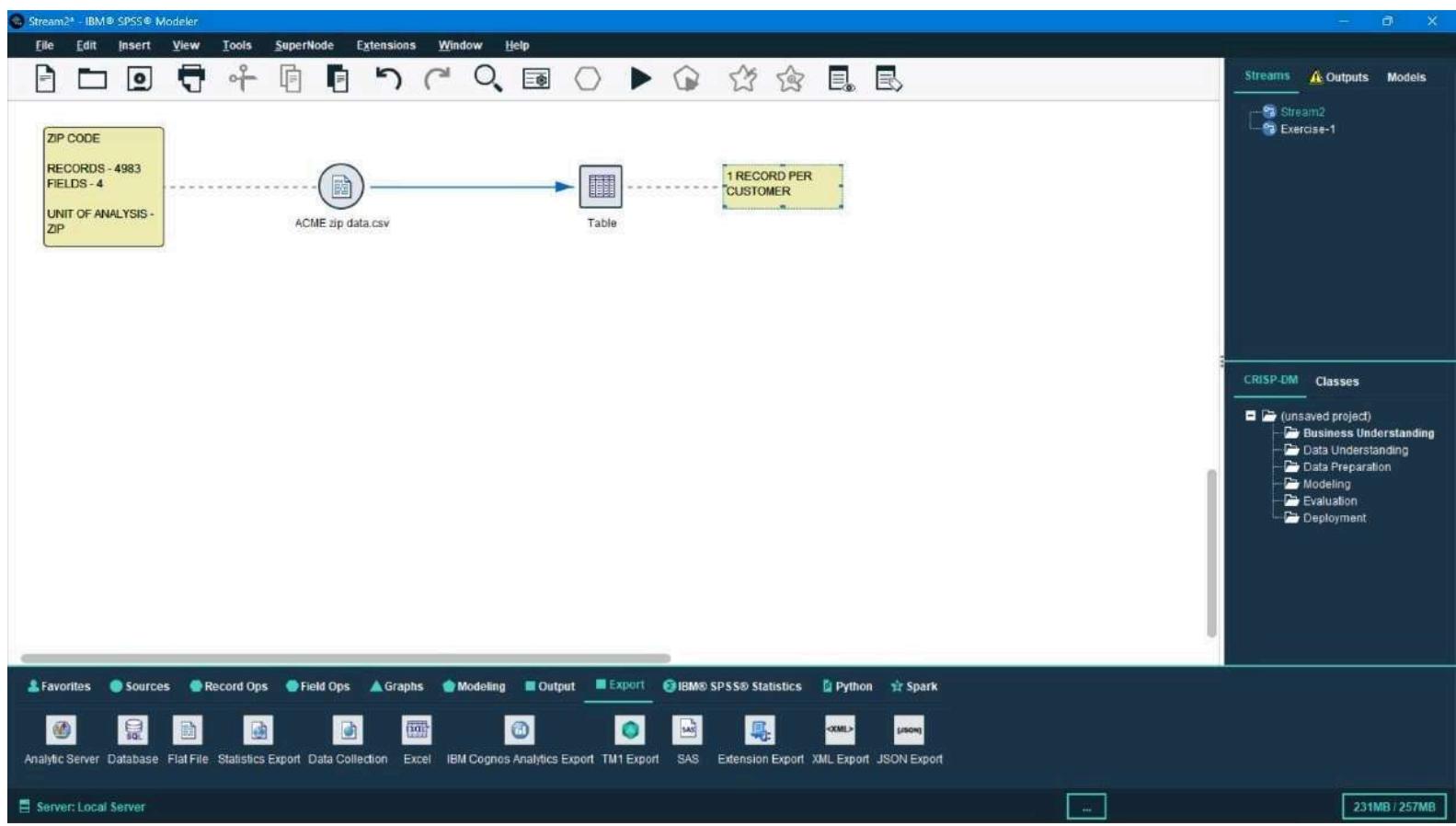


Table (4 fields, 4,983 records)

File Edit Generate

	ZIP	pct_working	pct_house_owners	pct_with_car
1	1001EZ	76	60	87
2	1001FA	76	60	88
3	1001GB	70	48	88
4	1001PJ	76	59	90
5	1001PR	86	69	91
6	1002AJ	88	72	92
7	1002EH	75	59	92
8	1002GZ	79	67	91
9	1002JC	76	57	87
10	1002JX	80	64	92
11	1002MP	83	67	91
12	1002W	83	70	92
13	1002WX	80	67	92
14	1002YB	73	57	93
15	1003X	71	56	89
16	1003OH	80	62	90
17	1003ON	88	74	90
18	1003TK	91	72	92
19	1003UB	67	49	93
20	1004CR	80	60	88
21	1004IB	81	65	89
22	1004PO	84	70	94
23	1004KV	85	70	88
24	1005GB	81	64	87
25	1005GR	77	62	88
26	1005SZ	70	55	95
27	1006FL	84	71	94
28	1006KL	70	54	96
29	1006VP	75	63	90
30	1006WP	75	61	92
31	1007AE	74	58	85
32	1007HU	76	63	86
33	1007JU	82	65	86
34	1007TN	83	70	90
35	1007VV	79	64	90
36	1008CD	75	60	89
37	1008DS	82	66	90
38	1008GT	71	61	92
39	1008IP	78	65	90
40	1008KU	82	66	96
41	1008LJ	82	66	91
42	1008NV	72	57	91
43	1008YE	84	69	89
44	1008YS	81	64	89

OK

PROJECT-3

Understanding your Data

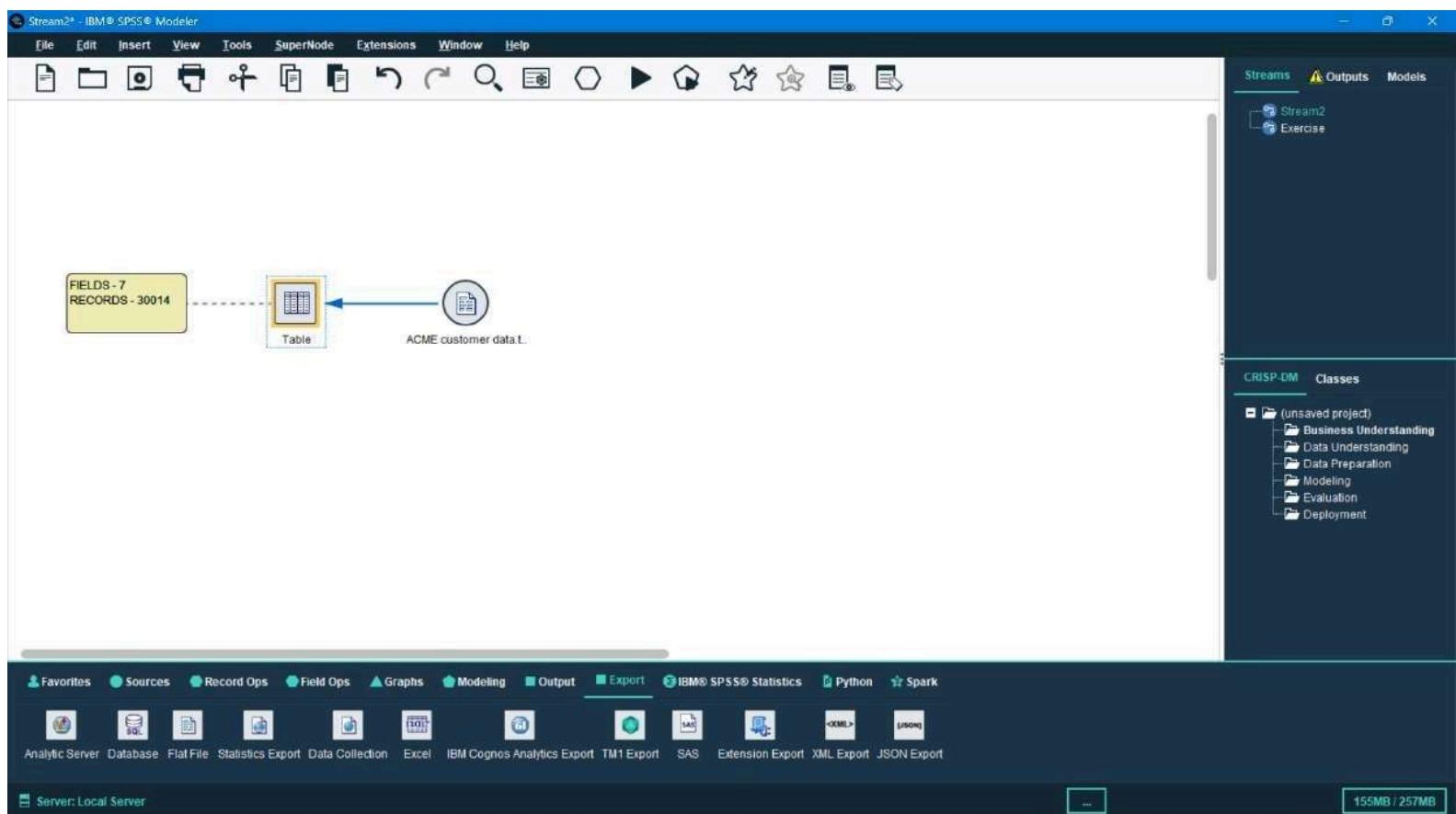


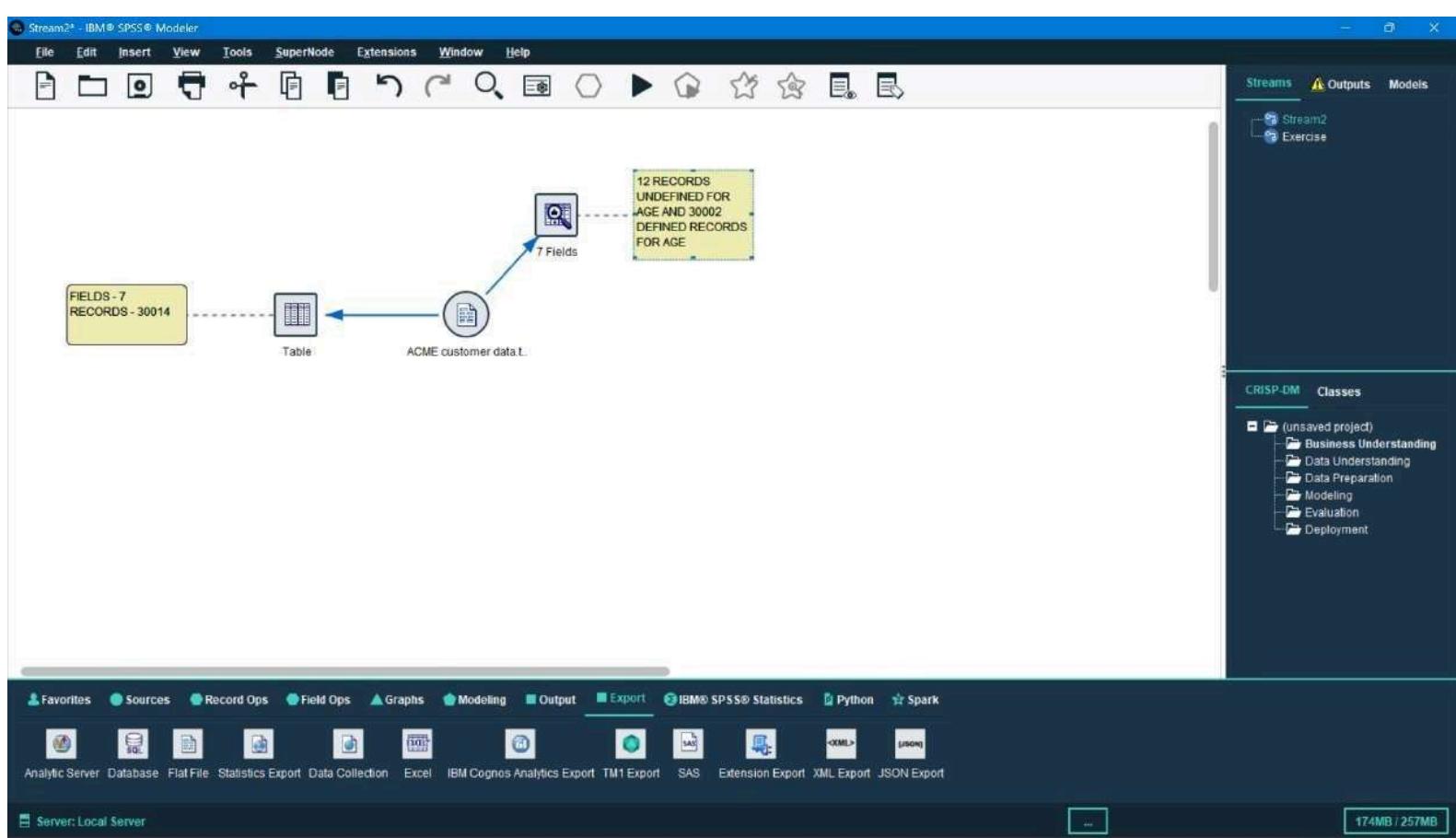
Table (7 fields, 30,014 records)

File Edit Generate Search

Table Annotations

	CUSTOMER_ID	GENDER	AGE	CREDITLIMIT	ZODIAC	E-MAIL ADDRESS	ZIP
1	723 F	20	9025	1	name7502@tnet.fr	1818BO	
2	724 M	28	5223	3	name25485@wwmail.org	1132DG	
3	725 F	0	658	8	name15543@wwmail.de	1803YT	
4	726 M	18	8302	6	name28335@zigzag.be	1205WR	
5	727 F	Sn...	997	8	name5354@tnet.jp	1711ON	
6	728	Sn...	\$null\$	\$null\$			
7	729 F	28	5898	5	name20536@wwmail.es	1254MR	
8	730 F	26	8488	11	name10414@tnet.inc	1723DG	
9	730 F	28	8488	11	name10414@tnet.inc	1723DG	
10	731 F	17	5006	2	name23372@wwmail.inc	1713AQ	
11	732 F	22	929	5	name20935@wwmail.es	1264EC	
12	733 F	26	3303	8	name5356@tnet.jp	1648BT	
13	734 F	28	8934	3	name17582@wwmail.de	1285VW	
14	735 F	27	9808	7	name5388@tnet.fr	1282NB	
15	736 M	31	4723	6	name10409@tnet.inc	1799IT	
16	737 M	28	672	7	name13849@tnet.uk	1802UO	
17	738	Sn...	\$null\$	\$null\$			
18	739 M	17	5984	5	name13848@tnet.uk	1361RL	
19	740 F	24	2621	9	name23365@wwmail.in	1164VN	
20	741 M	34	8062	6	name3188@molbe.cat	1767YN	
21	742 M	27	9797	4	name1606@omejor.es	1681HP	
22	743 F	23	2687	6	name29056@zigzag.cat	1287PI	
23	744 M	27	128	9	name28323@zigzag.be	1023AR	
24	745 F	28	182	11	name20530@wwmail.es	1737ZI	
25	746 M	26	2301	9	name13847@tnet.uk	1625TZ	
26	747 F	19	2704	8	name10401@tnet.inc	1684YE	
27	748 M	26	4033	7	name13846@tnet.uk	1508DU	
28	749	Sn...	\$null\$	\$null\$			
29	750 F	30	9044	6	name26894@wwmail.uk	1228IK	
30	751 F	19	1975	12	name6386@tnet.fr	1567CK	
31	752 F	23	3619	9	name1609@omejor.es	1315FI	
32	753 M	21	3517	8	name29727@zigzag.com	1753OI	
33	754 M	19	701	8	name29856@zigzag.es	1229DC	
34	755 M	22	1878	3	name17578@wwmail.de	1036LJ	
35	756 M	17	4023	10	name29046@zigzag.cat	1002W	
36	757 M	25	4946	10	name10389@tnet.inc	1799KZ	
37	758 F	30	9946	8	name5354@tnet.fr	1784IE	
38	759 F	35	1522	6	name3172@molbe.cat	1183FC	
39	760 F	23	7458	2	name43@hmail.de	1492WX	
40	761 F	25	1340	10	name5357@tnet.jp	1569VY	
41	761 F	29	1340	10	name5357@tnet.jp	1569VY	
42	761 F	33	1340	10	name5357@tnet.jp	1569VY	
43	762 F	25	7136	3	name255@hmail.inc	1558OG	
44	763 F	37	1723	4	name25457@wwmail.on	1582SR	

OK



Data Audit of [7 fields] #1

File Edit Generate

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
CUSTOMER...		Continuous	723	30722	15723.123	8661.140	-0.000	--	30014
GENDER		Categorical	--	--	--	--	--	5	30008
AGE		Continuous	0	999	27.129	44.055	21.359	--	30002
CREDITLIMIT		Continuous	-999999	9998	3599.487	32922.992	-30.208	--	30008
ZODIAC		Continuous	1	12	6.502	3.195	-0.008	--	30008
E-MAIL ADD...		Categorical	--	--	--	--	--	--	30007
ZIP		Categorical	--	--	--	--	--	--	30008

*Indicates a multimode result. ^ Indicates a sampled result.

OK

Stream2 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream2 Exercise

CRISP-DM Classes

(unsaved project)

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 223MB / 257MB

Stream2 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream2 Exercise

CRISP-DM Classes

(unsaved project)

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

AGE Values

Measurement: Continuous Storage: Integer Depth: 0 Model Field...

Values:

Lower: 15 Upper: 75

Labels...

Check values: Warn

Define blanks

Missing values

Range: to: Null White space

Description:

OK Cancel Help

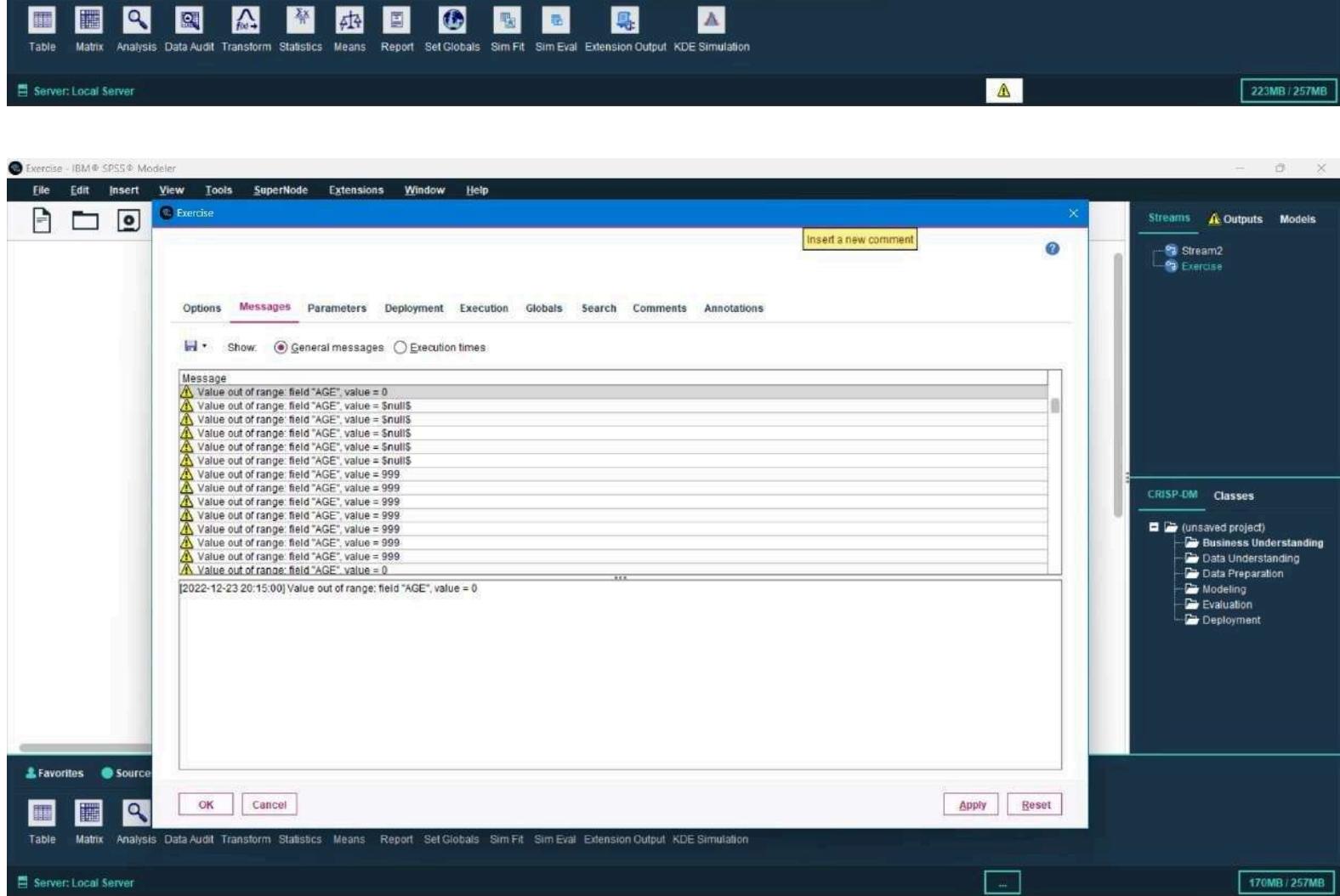
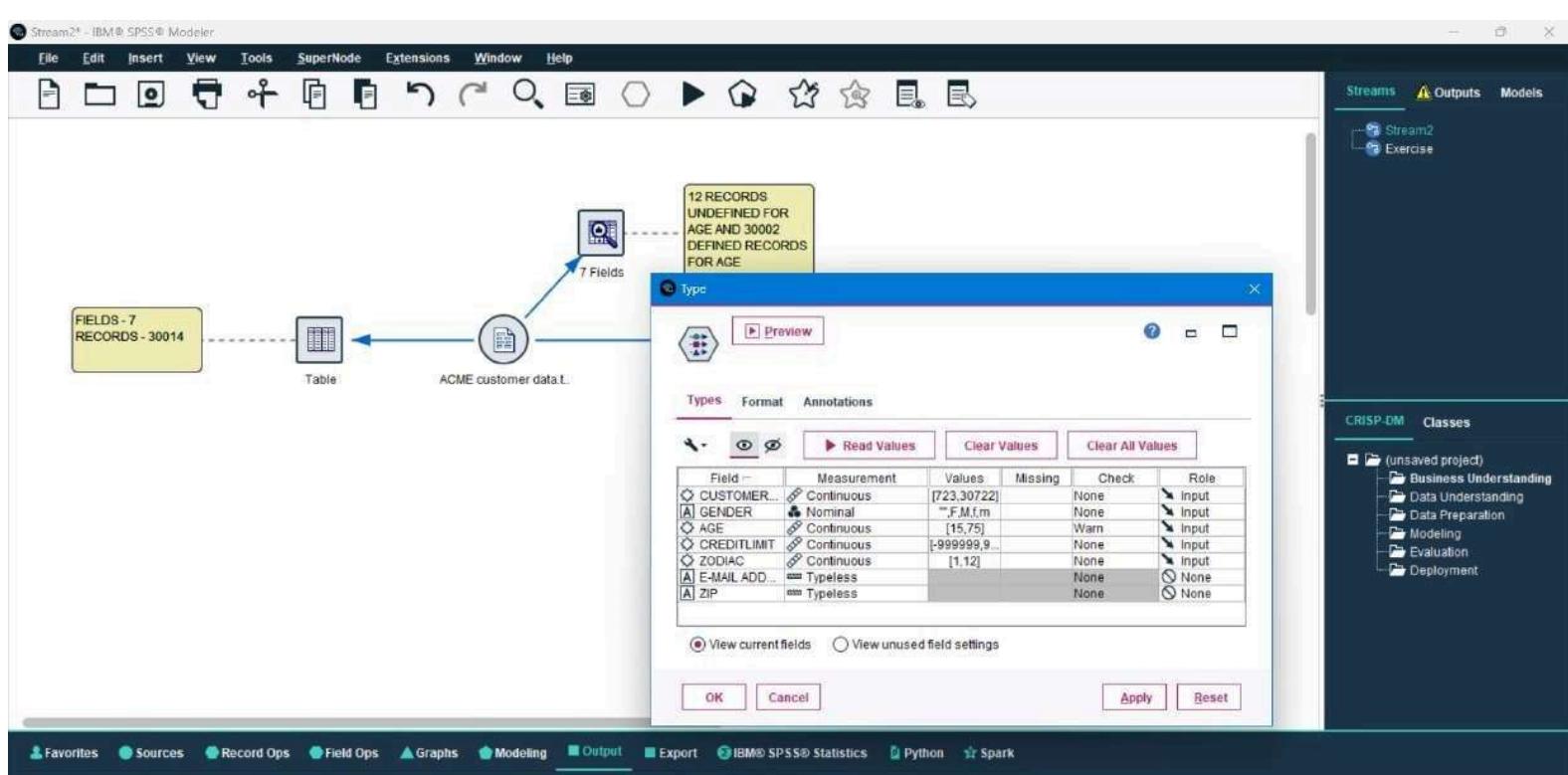
View current items View unused field settings

OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 223MB / 257MB



Stream2 - IBM® SPSS® Modeler

The screenshot shows a data flow in the Stream2 interface. A 'Table' node (labeled 'ACME customer data.t') has an arrow pointing to a 'Type' node. The 'Type' node has an arrow pointing to a 'AGE Values' dialog box. The dialog box contains the following information:

- 12 RECORDS UNDEFINED FOR AGE AND 30002**
- 7 Fields**
- AGE Values** (Measurement: Continuous, Storage: Integer, Depth: 0)
- Values:** Specified values and labels (Lower: 15, Upper: 75)
- Labels...**
- Check values:** Warn
- Define blanks** (checkbox checked)
- Missing values:** 0, 999
- Range:** (checkbox unchecked)
- Null:** (checkbox checked)
- White space:** (checkbox checked)
- Description:** (empty)

At the bottom of the dialog are 'OK', 'Cancel', and 'Help' buttons.

On the left, there's a yellow box labeled 'FIELDS - 7 RECORDS - 30014'. On the right, the CRISP-DM Classes sidebar shows an 'Exercise' project with various stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

Stream2 - IBM® SPSS® Modeler

The screenshot shows a modified data flow. The 'Table' node (labeled 'ACME customer data.t') has an arrow pointing to a 'Type' node. The 'Type' node has two arrows: one pointing to a 'Table' node and another pointing to a 'MEAN AGE - 25.145' node. The 'Type' node also has an arrow pointing to a yellow box containing the following information:

- 12 RECORDS UNDEFINED FOR AGE AND 30002**
- DEFINED RECORDS FOR AGE**
- Type**
- FOR AGE**
- LOWER RANGE - 15**
- UPPER RANGE - 75**
- CHECK VALUE - ABORT**
- MISSING VALUES - 0 and 999**

On the left, there's a yellow box labeled 'FIELDS - 7 RECORDS - 30014'. On the right, the CRISP-DM Classes sidebar shows an 'Exercise' project with various stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

Data Audit of [CUSTOMER_ID GENDER AGE CREDITLIMIT ZODIAC]

Audit **Quality** **Annotations**

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
CUSTOMER_ID		Continuous	723	30722	15723.123	8661.140	-0.000	-	30014
GENDER		Nominal	-	-	-	-	-	5	30008
AGE		Continuous	15	47	25.145	4.789	0.240	-	29867
CREDITLIMIT		Continuous	-999999	9998	3599.487	32922.992	-30.208	-	30008
ZODIAC		Continuous	1	12	6.502	3.195	-0.008	-	30008

Histogram of AGE

Graph Annotations

OK

* Indicates a multimodal result. ^ Indicates a sampled result

OK

PROJECT-4

Setting the Unit of Analysis

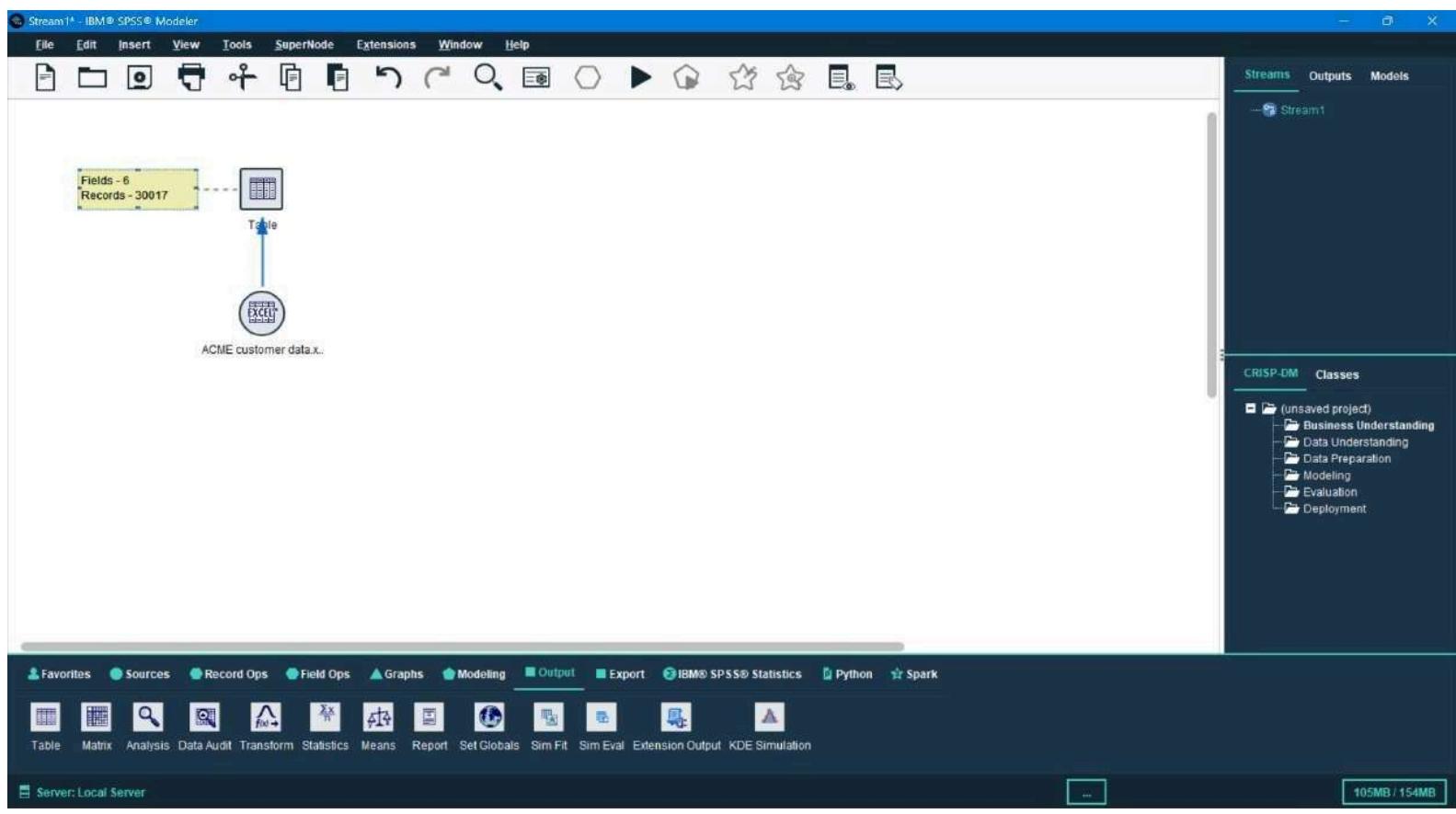


Table (6 fields, 30,017 records) #2

File Edit Generate ?

Table Annotations

	CUSTOMER_ID	GENDER	CREDITLIMIT	ZODIAC	E-MAIL	ADDRESS	ZIP
1	723 000 F	9026 000	1.000	name7502@netfr	1818BQ		
2	724 000 M	5223 000	3.000	name25485@wwmail.org	1132DG		
3	725 000 F	668 000	8.000	name15543@wwmail.de	1803YT		
4	726 000 M	8302 000	5.000	name28335@zigzag.be	1205WR		
5	727 000 F	997 000	8.000	name5354@net.jp	1711ON		
6	728 000 M	1955 000	3.000	name20637@wwmail.es	1055FG		
7	729 000 F	6898 000	5.000	name20636@wwmail.es	1254MR		
8	730 000 F	6488 000	11.000	name10414@net.inc	1723DG		
9	730 000 F	6488 000	11.000	name10414@net.inc	1723DG		
10	730 000 F	6488 000	11.000	name10414@net.inc	1723DG		
11	731 000 F	5008 000	2.000	name23372@wwmail.inc	1713AQ		
12	732 000 F	929 000	5.000	name20635@wwmail.es	1264EC		
13	733 000 F	3303 000	8.000	name5355@net.jp	1648BT		
14	734 000 F	8934 000	3.000	name17582@wwmail.de	1285XV		
15	735 000 F	9808 000	7.000	name6388@netfr	1282NB		
16	736 000 M	4723 000	6.000	name10409@net.inc	1799IT		
17	737 000 M	672 000	7.000	name13849@net.uk	1802UO		
18	738 000 F	7801 000	8.000	name25473@wwmail.org	1971NK		
19	739 000 M	5984 000	5.000	name13848@net.uk	1361RL		
20	740 000 F	2621 000	9.000	name23396@wwmail.inc	1164VN		
21	741 000 M	6062 000	6.000	name3188@molibe.cat	1767YN		
22	742 000 M	9797 000	4.000	name1605@lomejor.es	1681HP		
23	743 000 F	2687 000	6.000	name29056@zigzag.cat	1287PI		
24	744 000 M	128 000	9.000	name28323@zigzag.be	1023AR		
25	745 000 F	182 000	11.000	name20630@wwmail.es	1737ZI		
26	746 000 M	2301 000	9.000	name13847@net.uk	1625T2		
27	747 000 F	2704 000	8.000	name10401@net.inc	1684YE		
28	748 000 M	4033 000	7.000	name13846@net.uk	1508DU		
29	749 000 M	1818 000	9.000	name17579@wwmail.de	1143VC		
30	750 000 F	9044 000	6.000	name26894@wwmail.uk	1228K		
31	751 000 F	1975 000	12.000	name6365@netfr	1567CX		
32	752 000 F	3619 000	9.000	name1609@lomejor.es	1315FI		
33	753 000 M	3517 000	8.000	name29727@zigzag.com	1753OI		
34	754 000 M	701 000	8.000	name29656@zigzag.es	1229DC		
35	755 000 M	1878 000	3.000	name17578@wwmail.de	1036LJ		
36	756 000 M	4023 000	10.000	name29046@zigzag.cat	1002VV		
37	757 000 M	4946 000	10.000	name10389@net.inc	1979KZ		
38	758 000 F	9946 000	8.000	name6354@netfr	1784IE		
39	759 000 F	1522 000	6.000	name3172@molibe.cat	1163FC		
40	760 000 F	7458 000	2.000	name43@hmail.de	1492WX		
41	761 000 F	1340 000	10.000	name5357@net.jp	1569YY		
42	761 000 F	1340 000	10.000	name5357@net.jp	1569YY		
43	761 000 F	1340 000	10.000	name5357@net.jp	1569YY		
44	762 000 F	7136 000	3.000	name255@hmail.inc	1568DZ		

OK

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models Stream1

ACME customer data.x.

Fields - 6 Records - 30017

Table

EXCEL

Distinct

Settings Composite Optimization Annotations

Mode: Include only the first record in each group

Key fields for grouping:

- CUSTOMER_ID
- GENDER
- CREDITLIMIT

Sort Order

Within groups, sort records by:

Field	Order
-------	-------

Default sort order: Ascending Descending

If no sort fields are selected then the order of records within each group is undefined

OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 109MB / 154MB

Table (6 fields, 30,000 records) #1

File Edit Generate

Table Annotations

	CUSTOMER_ID	GENDER	CREDITLIMIT	ZODIAC	E-MAIL ADDRESS	ZIP
1	723.000	F	9026.000	1.000	name7502@tnet.fr	181880
2	724.000	M	5223.000	3.000	name25485@wwwmail.org	11320G
3	725.000	F	668.000	8.000	name15543@wwwmail.de	1803YT
4	726.000	M	8302.000	5.000	name28335@zigzag.be	1205WR
5	727.000	F	997.000	8.000	name5356@tnet.jp	1711ON
6	728.000	M	1955.000	3.000	name20637@wwwmail.es	1055FQ
7	729.000	F	6899.000	5.000	name20636@wwwmail.es	1254MR
8	730.000	F	6488.000	11.000	name10414@tnet.ln	1723DG
9	731.000	F	5006.000	2.000	name23372@wwwmail.inc	1713AQ
10	732.000	F	929.000	5.000	name20635@wwwmail.es	1264EC
11	733.000	F	3303.000	8.000	name5355@tnet.jp	1648BT
12	734.000	F	8934.000	3.000	name17582@wwwmail.de	1285XV
13	735.000	F	9808.000	7.000	name6389@tnet.fr	1282NB
14	736.000	M	4723.000	6.000	name10409@tnet.inc	1799IT
15	737.000	M	672.000	7.000	name13849@tnet.uk	1802UO
16	738.000	F	7801.000	8.000	name25473@wwwmail.org	1971NK
17	739.000	M	5984.000	5.000	name13848@tnet.uk	1361RL
18	740.000	F	2621.000	9.000	name23356@wwwmail.inc	1164VN
19	741.000	M	6062.000	8.000	name3188@molbe.cat	1767YN
20	742.000	M	9797.000	4.000	name1505@lomejor.es	1681HP
21	743.000	F	2687.000	6.000	name29056@gzag.cat	1287PI
22	744.000	M	128.000	9.000	name28323@gzag.cat	1023AR
23	745.000	F	182.000	11.000	name20630@wwwmail.es	1737ZI
24	746.000	M	2301.000	9.000	name13847@tnet.uk	1625Tz
25	747.000	F	2704.000	8.000	name10401@tnet.ln	1684YE
26	748.000	M	4033.000	7.000	name13846@tnet.uk	1508DU
27	749.000	M	1818.000	9.000	name17579@wwwmail.de	1143VC
28	750.000	F	9044.000	5.000	name20834@wwwmail.uk	1228IK
29	751.000	F	1975.000	12.000	name6356@tnet.fr	1567CK
30	752.000	F	3819.000	9.000	name1809@lomejor.es	1315FI
31	753.000	M	3517.000	8.000	name29727@gzag.com	1753OI
32	754.000	M	701.000	8.000	name29856@gzag.es	1229DC
33	755.000	M	1878.000	3.000	name17578@wwwmail.de	1036LJ
34	756.000	M	4023.000	10.000	name29046@gzag.cat	1002WV
35	757.000	M	4946.000	10.000	name10389@tnet.inc	1979KZ
36	758.000	F	9946.000	8.000	name6354@tnet.fr	1784IE
37	759.000	F	1522.000	6.000	name3172@molbe.cat	1183FC
38	760.000	F	7458.000	2.000	name43@wwwmail.de	1492WX
39	761.000	F	1340.000	10.000	name5357@tnet.jp	1569YY
40	762.000	F	7135.000	3.000	name255@tnet.inc	1558OG
41	763.000	F	1723.000	4.000	name25457@wwwmail.org	1588SB
42	764.000	F	3799.000	7.000	name1615@lomejor.es	1972BQ
43	765.000	M	1326.000	2.000	name5359@tnet.jp	1109VD
44	766.000	F	8432.000	2.000	name17576@wwwmail.de	1034PF

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Fields - 6 Records - 30017

Table ACME customer data.x. Distinct Table Fields - 6 Records - 30000
7 duplicate records are removed

CRISP-DM Classes

(unsaved project) Business Understanding Data Understanding Data Preparation Modeling Evaluation Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 128MB / 154MB

```
graph LR; A[ACME customer data.x.] --> B{Distinct}; B --> C[Table]; C --> D[Fields - 6  
Records - 30000  
7 duplicate records  
are removed]
```

Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Fields - 6 Records - 30017

Table ACME customer data.x. Distinct Table Fields - 6 Records - 30000
7 duplicate records are removed

Fields - 3 Records - 4018

Table ACME purchases 1999

CRISP-DM Classes

(unsaved project) Business Understanding Data Understanding Data Preparation Modeling Evaluation Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Select Sample Sort Balance Distinct Aggregate RFM Aggregate Merge Append Streaming TS SMOTE Extension Transform Space-Time-Boxes Streaming TCM CPLEX Optimization

Server: Local Server 136MB / 154MB

```
graph LR; A[ACME customer data.x.] --> B{Distinct}; B --> C[Table]; C --> D[Fields - 6  
Records - 30000  
7 duplicate records  
are removed]; E[ACME purchases 1999] --> F[Table]; F --> G[Fields - 3  
Records - 4018]
```

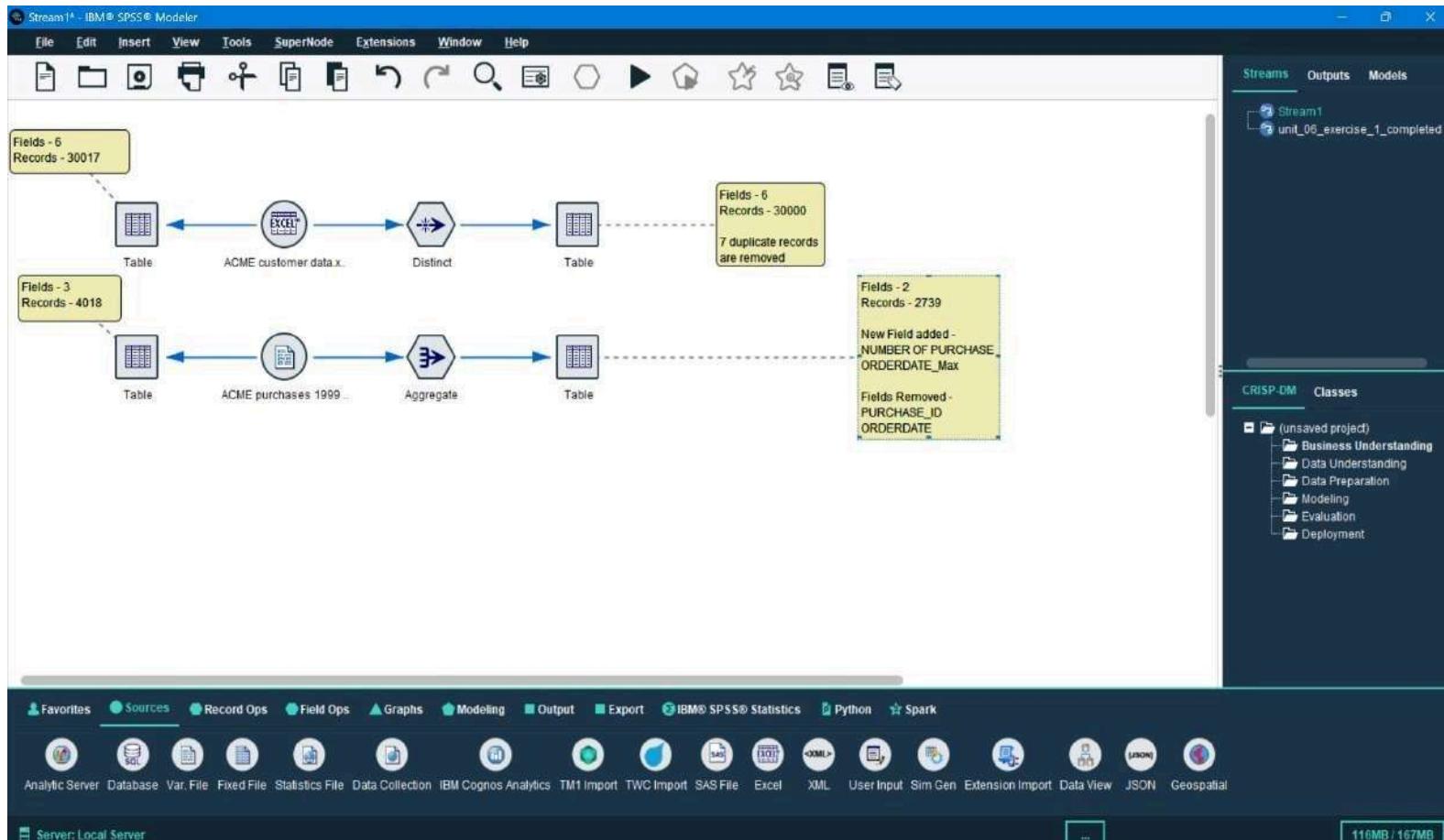
Table (3 fields, 4,018 records) #4

File Edit Generate Search

Table Annotations

	CUSTOMER_ID	PURCHASE_ID	ORDERDATE
1	724	8037 1999-07-11	
2	727	9324 1999-02-21	
3	728	18385 2004-05-22	
4	730	6966 2002-11-29	
5	731	8620 2000-07-21	
6	731	6336 2001-11-09	
7	731	7583 2003-07-17	
8	732	10648 2003-04-22	
9	732	8330 2003-05-10	
10	733	10618 2001-05-01	
11	733	5838 2003-12-01	
12	736	8296 2000-11-13	
13	736	6357 2003-10-22	
14	737	7453 2003-03-09	
15	738	6025 2002-11-13	
16	739	6253 2003-01-28	
17	740	8192 1999-12-07	
18	740	6902 2001-02-05	
19	741	9554 1999-07-15	
20	741	8143 2001-10-04	
21	742	5987 2000-01-22	
22	743	10245 2001-01-22	
23	743	7172 2003-09-16	
24	744	18386 2004-08-09	
25	745	8532 2001-08-24	
26	745	8765 2001-10-20	
27	748	7272 2000-01-20	
28	748	6540 2003-08-27	
29	751	10137 1999-01-09	
30	751	7209 1999-10-24	
31	752	6981 2000-04-10	
32	752	9565 2002-12-13	
33	753	6227 2002-01-18	
34	755	18387 2004-05-14	
35	756	8774 2004-04-16	
36	756	9721 2004-09-15	
37	756	18388 2004-11-15	
38	757	10683 2000-05-01	
39	759	8252 2001-08-15	
40	759	8330 2003-03-04	
41	760	10582 2001-11-19	
42	761	6492 1999-12-06	
43	761	6867 2002-04-18	
44	761	5905 2003-11-24	

OK



Stream1 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

Streams Outputs Models Stream1 unit_06_exercise_1_completed

Fields - 6 Records - 30017
Table ACME customer data.x.

Fields - 3 Records - 4018
Table ACME purchases 1999..

Aggregate

Preview

Settings Optimization Annotations

Key fields: CUSTOMER_ID

Basic Aggregates Aggregate fields:

Field	Sum	Mean	Min	Max	SDev	Median	Count	Variance	1st Quartile	3rd Quartile
ORDERDATE				<input checked="" type="checkbox"/>						

Default mode: Sum Mean Min Max SDev Median Count Variance 1st Quartile 3rd Quartile

New field name extension: Add as: Suffix Prefix

Include record count in field NUMBER OF PURCHASES

Aggregate Expressions

Field	Expression

OK Cancel Apply Reset

CRISP-DM Classes

(unsaved project)

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Graph

Analytic Server Database Var. File Fixed File Statistics File Data Collector

Server: Local Server 116MB / 167MB

Table (3 fields, 2,739 records) #2

File Edit Generate

Table Annotations

CUSTOMER_ID	ORDERDATE_Max	NUMBER OF PURCHASES
1	724 1999-07-11	1
2	727 1999-02-21	1
3	728 2004-05-22	1
4	730 2002-11-29	1
5	731 2003-07-17	3
6	732 2003-05-10	2
7	733 2003-12-01	2
8	736 2003-10-22	2
9	737 2003-03-09	1
10	738 2002-11-13	1
11	739 2003-01-28	1
12	740 2001-02-05	2
13	741 2001-10-04	2
14	742 2000-01-22	1
15	743 2003-09-16	2
16	744 2004-08-09	1
17	745 2001-10-20	2
18	748 2003-08-27	2
19	751 1999-10-24	2
20	752 2002-12-13	2
21	753 2002-01-18	1
22	755 2004-05-14	1
23	756 2004-11-15	3
24	757 2000-05-01	1
25	759 2003-03-04	2
26	760 2001-11-19	1
27	761 2002-11-24	3
28	762 2002-12-10	1
29	764 1999-09-07	1
30	765 2002-02-10	2
31	766 2004-06-07	3
32	770 2003-10-07	2
33	773 2001-08-24	2
34	775 2000-10-14	1
35	775 1999-01-26	1
36	777 1999-03-07	1
37	778 2001-10-15	1
38	779 2004-03-30	2
39	780 2003-03-07	1
40	781 2004-05-20	3
41	783 2002-05-26	1
42	784 2003-02-22	3
43	785 2000-10-03	1
44	788 2002-08-13	2

OK

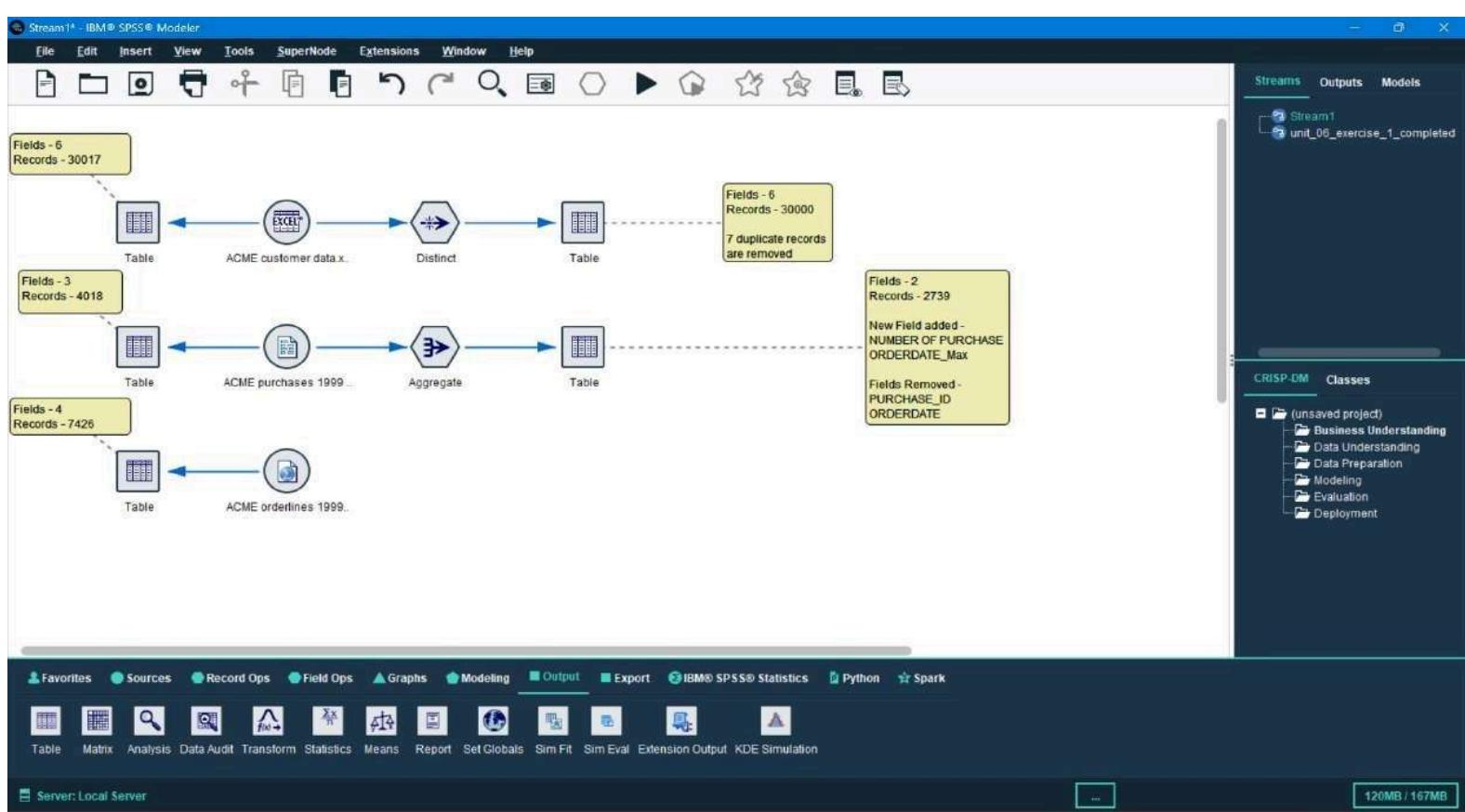


Table (4 fields, 7426 records) #2

File Edit Generate ...

	PURCHASE_ID	ITEM_ID	PRODUCT_ID	PRICE
1	5723.000	16243...	10726.000	197.040
2	5723.000	16843...	10736.000	143.280
3	5723.000	15183...	10746.000	129.840
4	5723.000	12442...	10748.000	478.320
5	5724.000	15244...	10732.000	114.840
6	5726.000	17840...	10725.000	217.770
7	5727.000	14843...	10723.000	103.480
8	5727.000	11494...	10733.000	164.100
9	5727.000	10942...	10742.000	198.550
10	5729.000	17131...	10723.000	243.480
11	5729.000	18058...	10730.000	279.060
12	5729.000	17320...	10732.000	462.980
13	5730.000	17384...	10738.000	672.390
14	5730.000	14449...	10752.000	262.960
15	5731.000	13120...	10731.000	43.940
16	5731.000	18201...	10736.000	702.900
17	5732.000	14202...	10731.000	45.520
18	5732.000	15940...	10736.000	311.040
19	5732.000	14529...	10747.000	260.960
20	5732.000	13400...	10748.000	322.920
21	5734.000	16808...	10730.000	460.000
22	5736.000	14193...	10724.000	296.950
23	5736.000	13599...	10747.000	308.460
24	5737.000	18262...	10723.000	413.350
25	5737.000	15915...	10725.000	486.900
26	5737.000	18351...	10731.000	299.320
27	5738.000	15948...	10734.000	298.370
28	5739.000	14598...	10742.000	30.450
29	5740.000	14587...	10737.000	201.680
30	5741.000	15278...	10741.000	319.200
31	5743.000	13011...	10747.000	30.570
32	5744.000	14427...	10744.000	75.520
33	5745.000	13764...	10747.000	60.560
34	5746.000	13768...	10729.000	389.000
35	5746.000	16540...	10738.000	112.280
36	5747.000	16269...	10737.000	296.700
37	5747.000	16460...	10752.000	228.080
38	5748.000	15546...	10736.000	274.400
39	5748.000	15610...	10739.000	803.840
40	5748.000	15387...	10745.000	165.060
41	5748.000	17958...	10751.000	35.970
42	5749.000	16579...	10740.000	288.630
43	5751.000	14353...	10724.000	542.400
44	5761.000	17425...	10749.000	202.440

OK

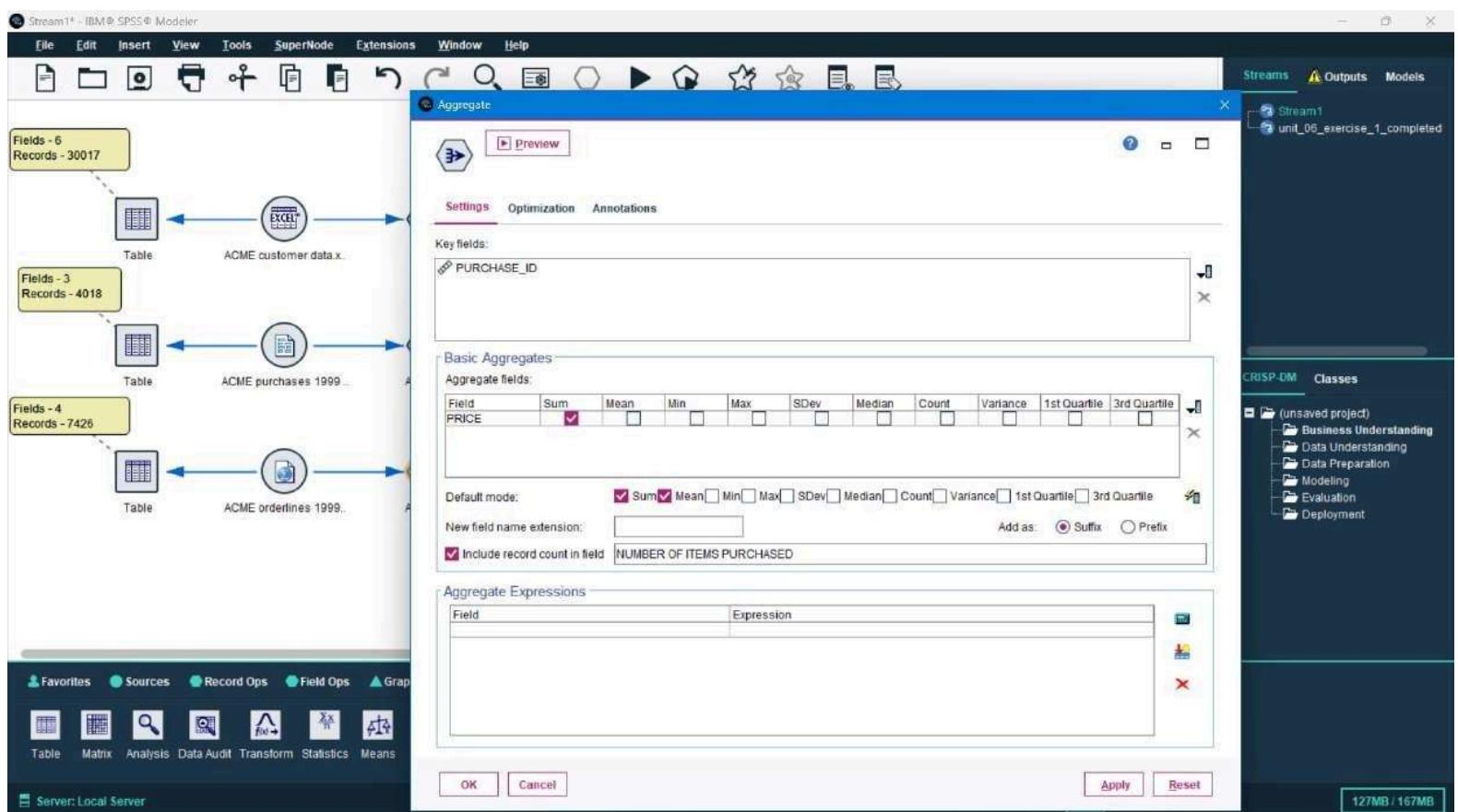
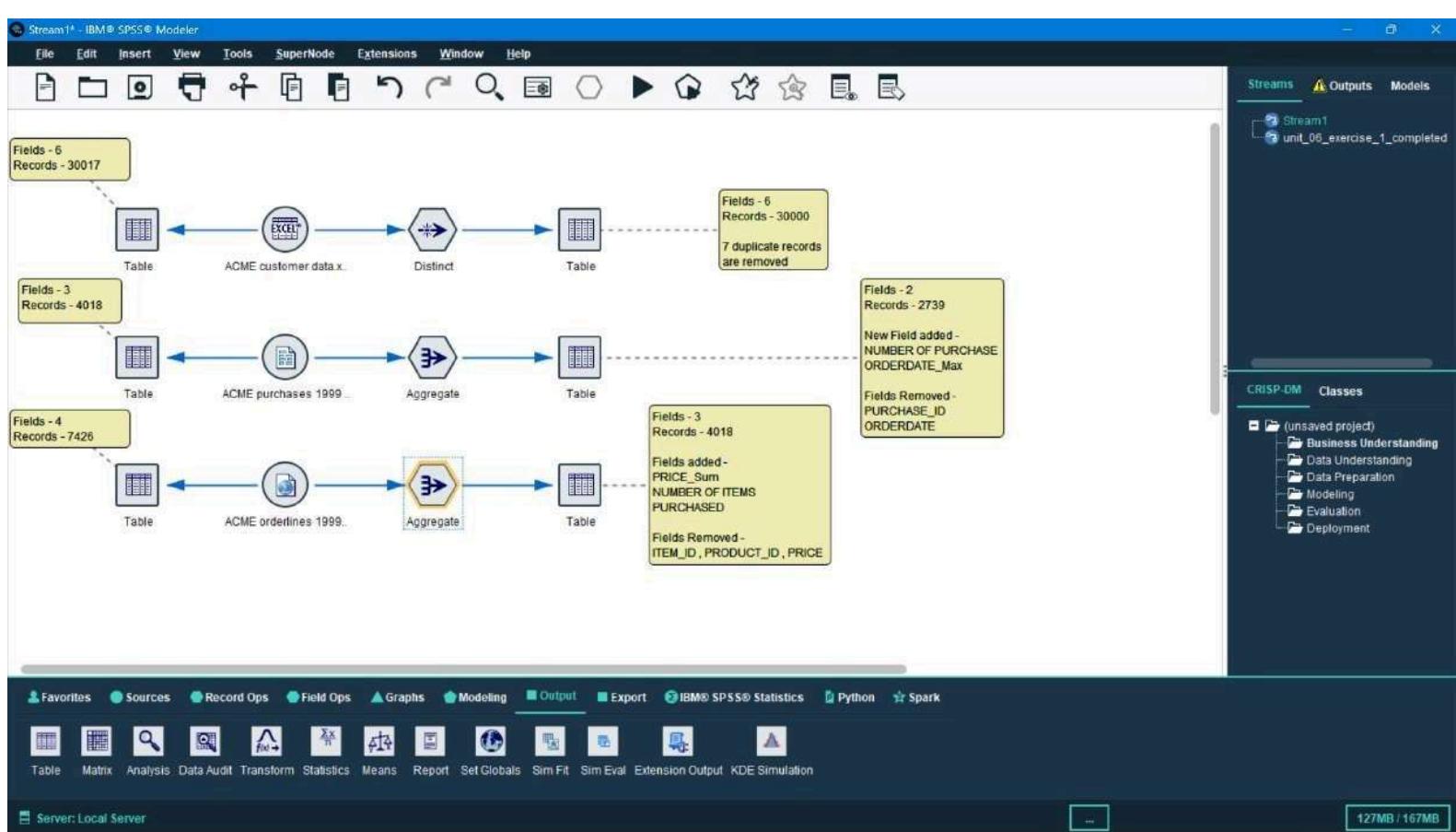


Table (3 fields, 4,018 records) #13

Annotations

PURCHASE_ID	PRICE_SUM	NUMBER OF ITEMS PURCHASED
1	5723.000	948.480
2	5724.000	114.840
3	5726.000	217.770
4	5727.000	466.130
5	5729.000	985.520
6	5730.000	935.350
7	5731.000	746.840
8	5732.000	940.540
9	5734.000	480.000
10	5736.000	605.410
11	5737.000	1199.570
12	5738.000	296.370
13	5739.000	30.450
14	5740.000	201.880
15	5741.000	319.200
16	5743.000	30.570
17	5744.000	75.520
18	5745.000	80.560
19	5746.000	501.280
20	5747.000	524.780
21	5748.000	1279.270
22	5749.000	288.630
23	5751.000	744.840
24	5753.000	886.360
25	5754.000	490.880
26	5755.000	31.620
27	5756.000	241.330
28	5758.000	141.570
29	5760.000	3.100
30	5761.000	1425.910
31	5762.000	300.300
32	5763.000	259.740
33	5765.000	950.460
34	5766.000	92.450
35	5767.000	136.040
36	5768.000	984.380
37	5769.000	840.800
38	5770.000	612.820
39	5772.000	350.000
40	5773.000	175.700
41	5774.000	445.020
42	5775.000	697.680
43	5776.000	228.830
44	5777.000	121.900

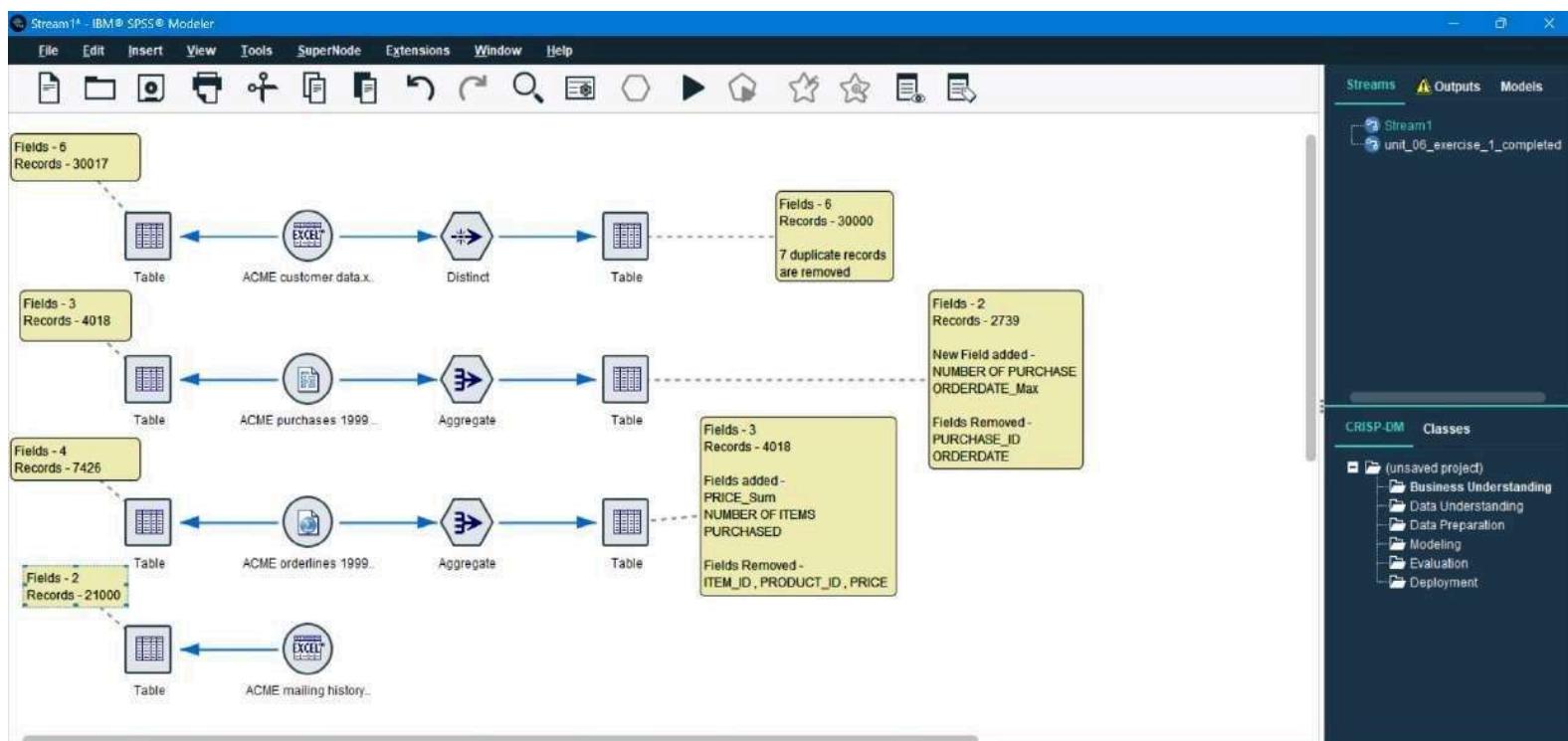
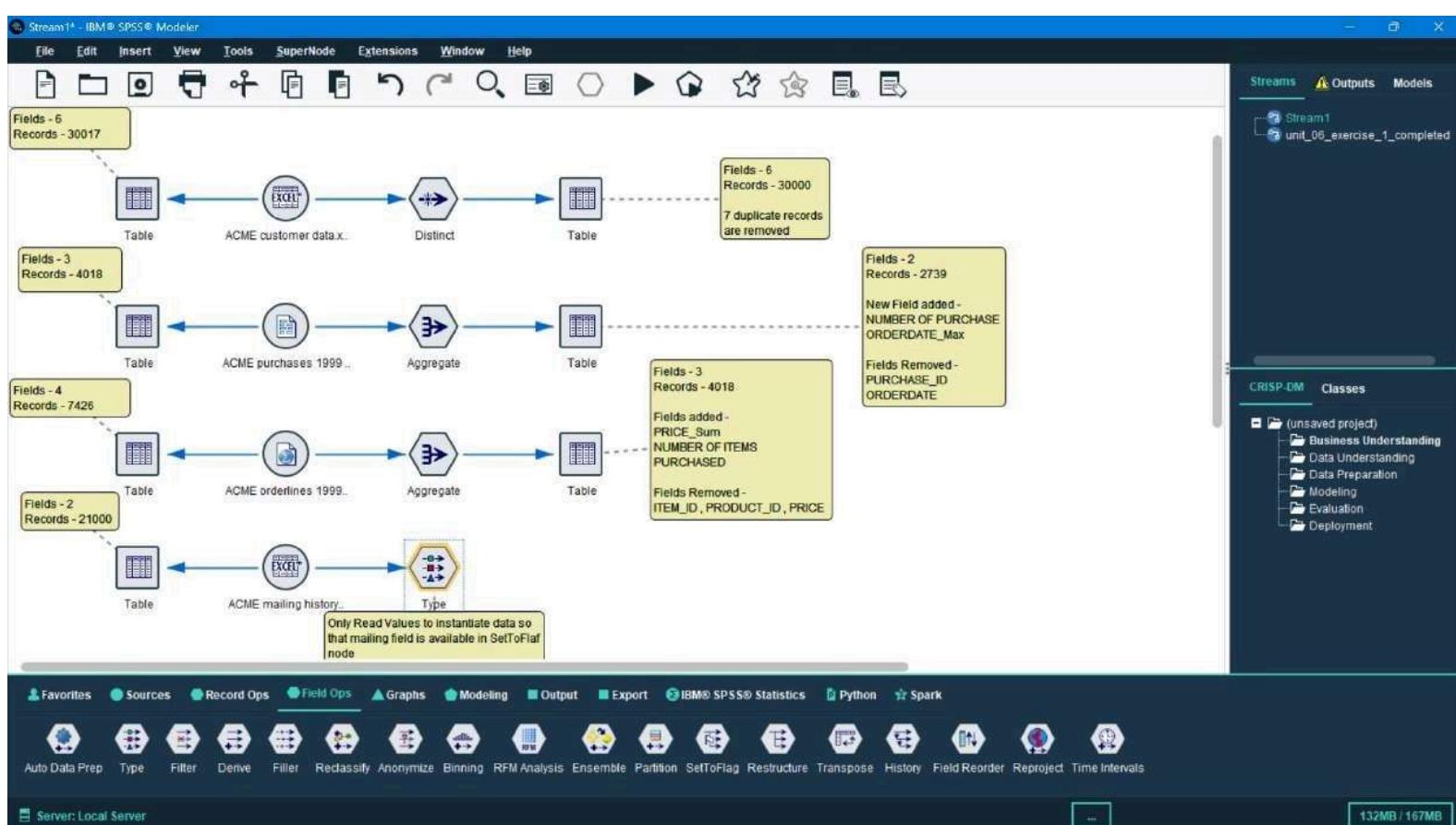
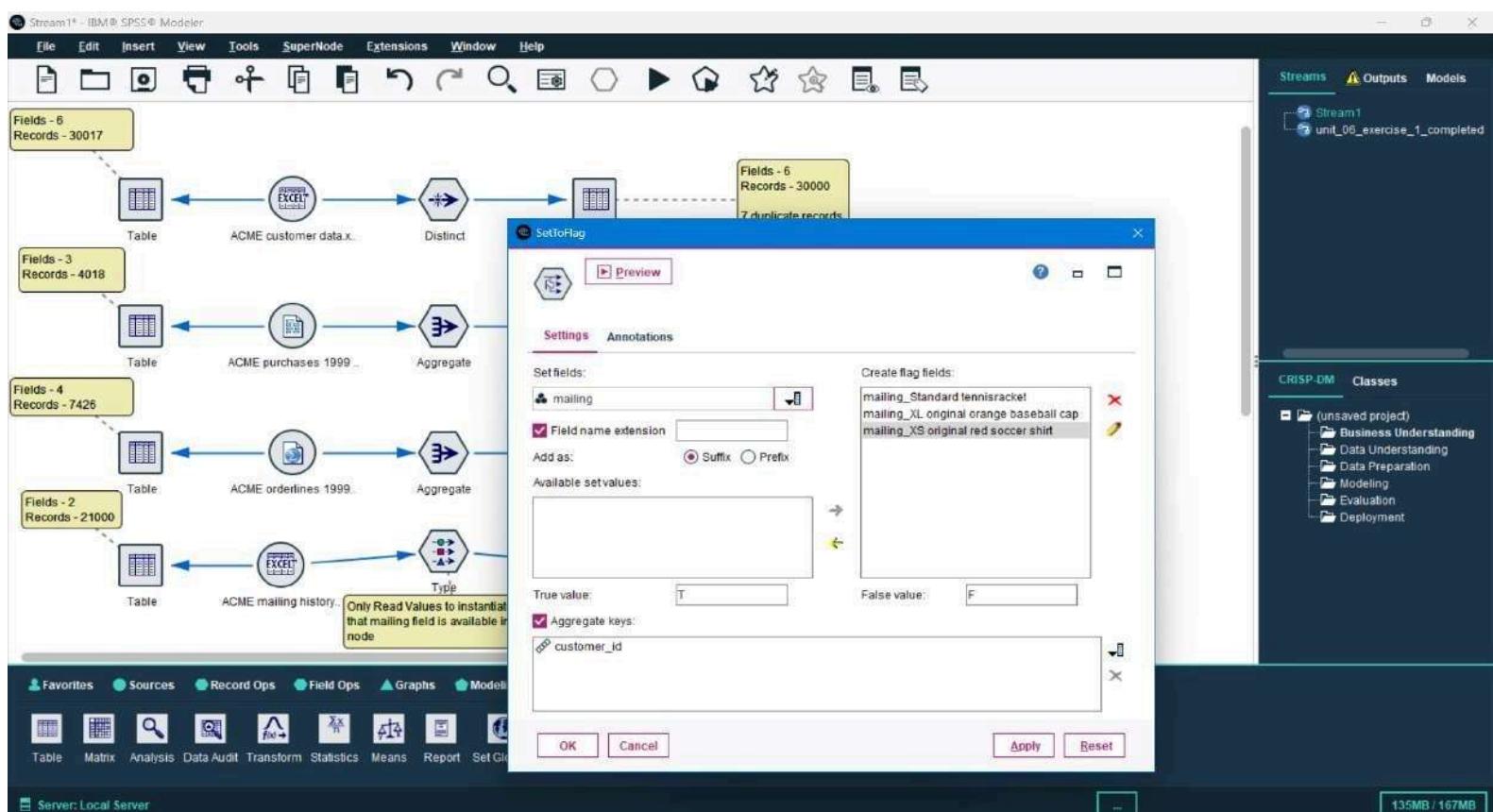
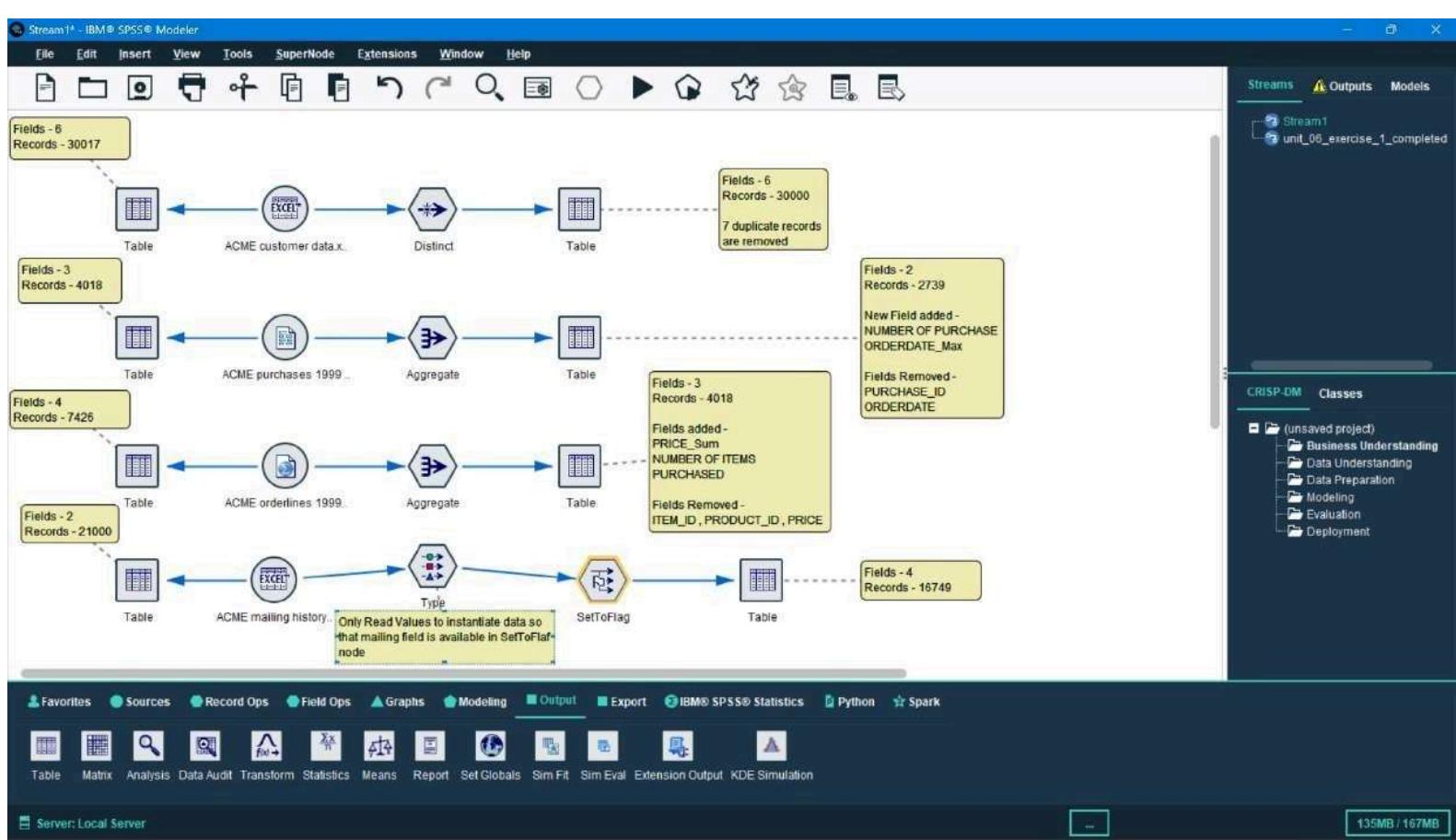


Table (2 Fields, 21,000 records) #1

Annotations

customer_id	mailing
1	723.000.XL original orange baseball cap
2	723.000.XS original red soccer shirt
3	724.000.XL original orange baseball cap
4	725.000.XL original orange baseball cap
5	726.000.XL original orange baseball cap
6	727.000 Standard tennisracket
7	727.000.XL original orange baseball cap
8	727.000.XS original red soccer shirt
9	728.000.XL original orange baseball cap
10	729.000.XL original orange baseball cap
11	730.000 Standard tennisracket
12	730.000.XL original orange baseball cap
13	731.000.XL original orange baseball cap
14	731.000.XS original red soccer shirt
15	732.000.XL original orange baseball cap
16	732.000.XS original red soccer shirt
17	733.000.XL original orange baseball cap
18	733.000.XS original red soccer shirt
19	734.000.XL original orange baseball cap
20	734.000.XS original red soccer shirt
21	735.000.XL original orange baseball cap
22	736.000.XL original orange baseball cap
23	736.000.XS original red soccer shirt
24	737.000 Standard tennisracket
25	737.000.XL original orange baseball cap
26	738.000.XL original orange baseball cap
27	739.000.XL original orange baseball cap
28	740.000 Standard tennisracket
29	740.000.XL original orange baseball cap
30	741.000.XL original orange baseball cap
31	742.000.XL original orange baseball cap
32	742.000.XS original red soccer shirt
33	743.000.XL original orange baseball cap
34	743.000.XS original red soccer shirt
35	744.000.XL original orange baseball cap
36	745.000.XL original orange baseball cap
37	746.000 Standard tennisracket
38	746.000.XL original orange baseball cap
39	747.000 Standard tennisracket
40	747.000.XL original orange baseball cap
41	748.000.XL original orange baseball cap
42	749.000.XL original orange baseball cap
43	750.000.XL original orange baseball cap
44	751.000.XL original orange baseball cap





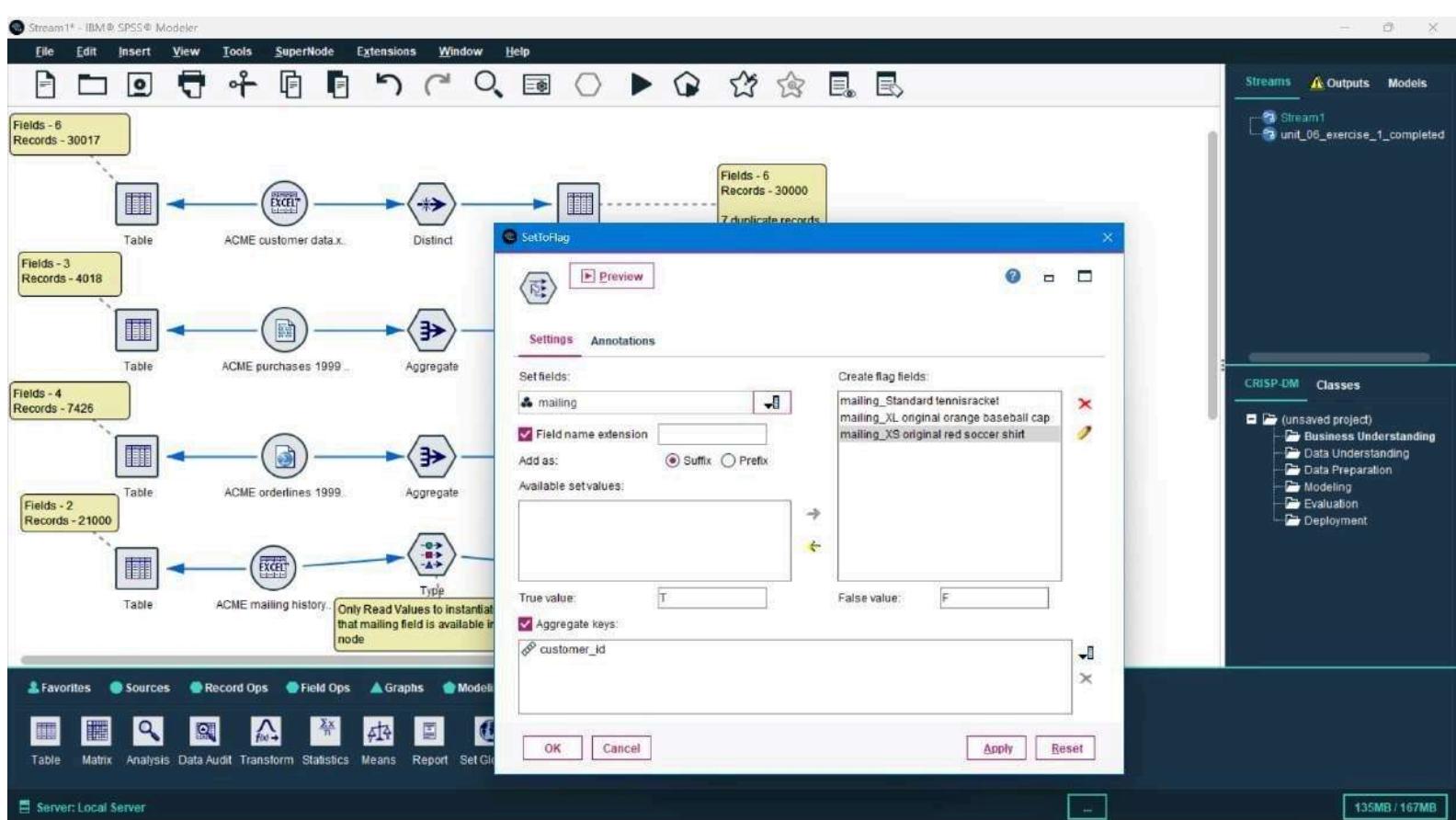


Table (4 Fields, 16,740 records) #2

File Edit Generate

Annotations			
customer_id	mailing_Standard tennisracket	mailing_XL original orange baseball cap	mailing_XS original red soccer shirt
1 723.000 F	T	T	
2 724.000 F	T	F	
3 725.000 F	T	F	
4 726.000 F	T	F	
5 727.000 T	T	T	
6 728.000 F	T	F	
7 729.000 F	T	F	
8 730.000 T	T	F	
9 731.000 F	T	T	
10 732.000 F	T	T	
11 733.000 F	T	T	
12 734.000 F	T	T	
13 735.000 F	T	F	
14 736.000 F	T	T	
15 737.000 T	T	F	
16 738.000 F	T	F	
17 739.000 F	T	F	
18 740.000 T	T	F	
19 741.000 F	T	F	
20 742.000 F	T	T	
21 743.000 F	T	T	
22 744.000 F	T	F	
23 745.000 F	T	F	
24 746.000 T	T	F	
25 747.000 T	T	F	
26 748.000 F	T	F	
27 749.000 F	T	F	
28 750.000 F	T	F	
29 751.000 F	T	T	
30 752.000 F	T	F	
31 753.000 F	T	T	
32 754.000 F	T	T	
33 755.000 F	T	F	
34 756.000 F	T	T	
35 757.000 T	T	F	
36 758.000 F	T	F	
37 759.000 F	T	F	
38 760.000 F	T	F	
39 761.000 F	T	F	
40 762.000 F	T	F	
41 763.000 F	T	F	
42 764.000 F	T	F	
43 765.000 F	T	T	
44 766.000 F	T	F	

OK

PROJECT-5

Integrating Data

Stream2* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

CRISP-DM Classes

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 149MB / 167MB

Stream2* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

CRISP-DM Classes

Preview from ACME purchases 1999 - 2004.dat Node (3 fields, 10... OK

	CUSTOMER_ID	PURCHASE_ID	ORDERD...
1	724	8037	1999-07-11
2	727	9324	1999-02-21
3	728	18385	2004-05-22
4	730	6966	2002-11-29
5	731	8620	2000-07-21
6	731	6336	2001-11-09
7	731	7583	2003-07-17
8	732	10648	2003-04-22
9	732	8330	2003-05-10
10	733	10618	2001-05-01

Preview from ACME purchases 2005 - 2010.dat Node (3 fields, 10... OK

	CUSTOMER_ID	PURCHASE_ID	ORDERD...
1	723	707867026	2008-03-01
2	723	707867024	2009-07-02
3	725	708270082	2005-09-14
4	725	708270077	2005-09-24
5	725	708270079	2006-11-08
6	725	708270086	2007-02-27
7	726	708477065	2005-01-14
8	726	708477054	2005-06-27
9	726	708477050	2007-12-26
10	726	708477059	2008-04-11

CRISP-DM Classes

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 150MB / 167MB

Stream2 - IBM® SPSS® Modeler

File Edit Insert Tools SuperNode Extensions Window Help

Streams Outputs Models

Fields - 3 Records - 71127

ACME purchases 1999

ACME purchases 2005

Append

Preview Append 2 datasets

No edits made

Inputs Append Annotations

Match fields by: Position Name Match case

Preview of field matches and structure

Output Field	1[ACME purchases 199...	2[ACME purchases 200...
CUSTOMER_ID	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
PURCHASE_ID	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
ORDERDATE	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Include fields from: Main dataset only All datasets

Tag records by including source dataset in field Input

OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 150MB / 167MB

Table (3 fields, 71,127 records) #3

File Edit Generate

Table Annotations

	CUSTOMER_ID	PURCHASE_ID	ORDERD...
1	724	8037	1999-07-11
2	727	9324	1999-02-21
3	728	18385	2004-05-22
4	730	6966	2002-11-29
5	731	8620	2000-07-21
6	731	6336	2001-11-09
7	731	7583	2003-07-17
8	732	10648	2003-04-22
9	732	8330	2003-05-10
10	733	10818	2001-05-01
11	733	5838	2003-12-01
12	736	8296	2000-11-13
13	736	6357	2003-10-22
14	737	7453	2003-03-09
15	738	6025	2002-11-13
16	739	6253	2003-01-28
17	740	8192	1999-12-07
18	740	6902	2001-02-05
19	741	9554	1999-07-15
20	741	8143	2001-10-04
21	742	5987	2000-01-22
22	743	10245	2001-01-22
23	743	7172	2003-09-16
24	744	18386	2004-08-09
25	745	8532	2001-08-24
26	745	8765	2001-10-20
27	748	7272	2000-01-20
28	748	6540	2003-08-27
29	751	10137	1999-01-09
30	751	7209	1999-10-24
31	752	6961	2000-04-10
32	752	9565	2002-12-13
33	753	8227	2002-01-18
34	755	18387	2004-05-14
35	756	8774	2004-04-16
36	756	9721	2004-09-15
37	756	18388	2004-11-15
38	757	10683	2000-05-01
39	759	8252	2001-08-15
40	759	9330	2003-03-04
41	760	10582	2001-11-19
42	761	6492	1999-12-06
43	761	6867	2002-04-18
44	761	5905	2002-11-24

OK

Stream2 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 145MB / 203MB

Stream2 - IBM® SPSS® Modeler

File Edit Preview from ACME orderlines 1999 - 2004.sav Node [4 fields, 10...]

File Edit Generate

	PURCHASE_ID	ITEM_ID	PRODUCT_ID	PRICE
1	5723	16243	10726	197.040
2	5723	16643	10736	143.280
3	5723	15183	10746	129.840
4	5723	12442	10748	478.320
5	5724	15244	10732	114.840
6	5726	17643	10725	217.770
7	5727	14843	10723	103.480
8	5727	11494	10733	164.100
9	5727	10942	10742	198.550
10	5729	17131	10723	243.480

OK

File Edit Generate

	PURCHASE_ID	ITEM_ID	PRODUCT_ID	PRICE
1	707867024	472754321	10735	69.521
2	707867024	5933029414	10741	152.308
3	707867026	3349835079	10734	324.903
4	708270077	5145284958	10734	73.191
5	708270077	4879332538	10739	452.667
6	708270079	4450722394	10724	256.858
7	708270082	5977515801	10723	271.570
8	708270082	2038221845	10742	263.599
9	708270082	7024719200	10745	244.274
10	708270082	2561070747	10746	90.724

OK

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 145MB / 203MB

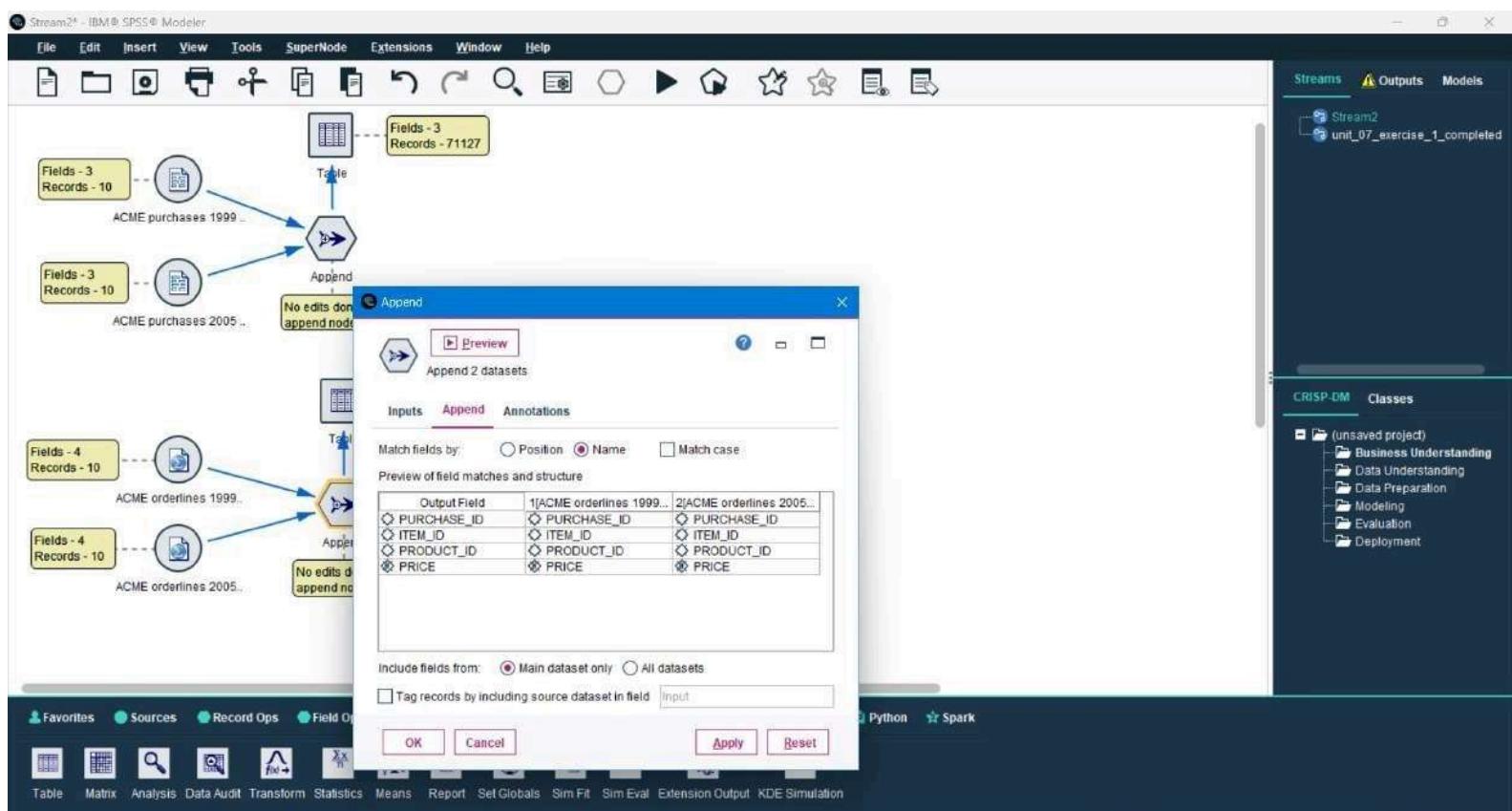


Table (4 fields, 169,734 records) #1

Annotations:

PURCHASE_ID	ITEM_ID	PRODUCT_ID	PRICE
1	5723	16243	10726
2	5723	16643	10736
3	5723	15183	10746
4	5723	12442	10748
5	5724	15244	10732
6	5726	17640	10725
7	5727	14843	10723
8	5727	11494	10733
9	5727	10942	10742
10	5729	17131	10723
11	5729	18058	10730
12	5729	17320	10732
13	5730	17384	10738
14	5730	14449	10752
15	5731	13120	10731
16	5731	18201	10736
17	5732	14202	10731
18	5732	15940	10736
19	5732	14529	10747
20	5732	13400	10748
21	5734	16808	10730
22	5736	14193	10724
23	5738	13599	10747
24	5737	18262	10723
25	5737	15915	10725
26	5737	18351	10731
27	5738	15948	10734
28	5739	14698	10742
29	5740	14587	10737
30	5741	15278	10741
31	5743	13011	10747
32	5744	14427	10744
33	5745	13764	10747
34	5746	13768	10729
35	5746	16540	10738
36	5747	16269	10737
37	5747	16460	10752
38	5748	15546	10736
39	5748	15610	10739
40	5748	15387	10745
41	5748	17858	10751
42	5749	15579	10740
43	5751	14353	10724
44	5751	17426	10749

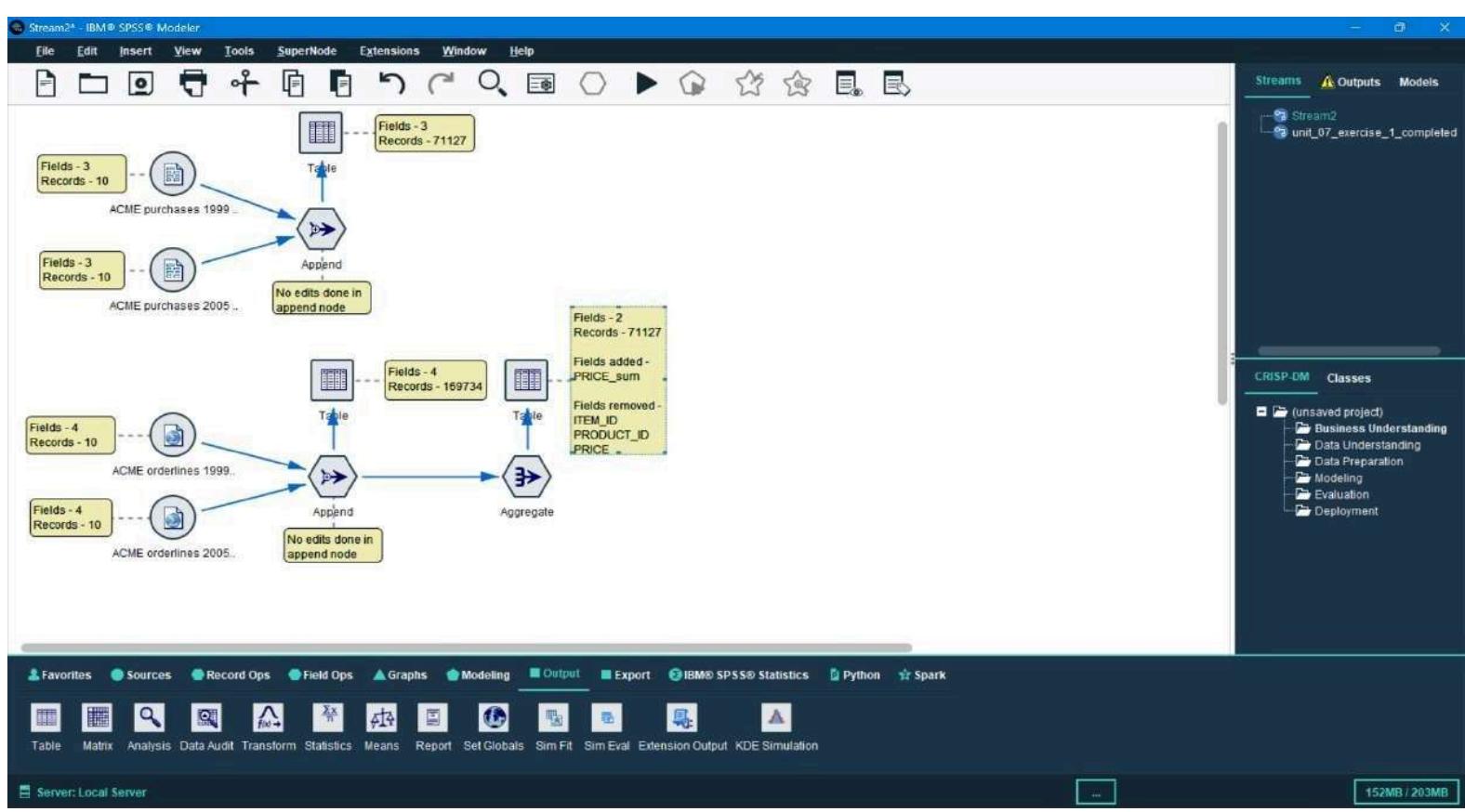


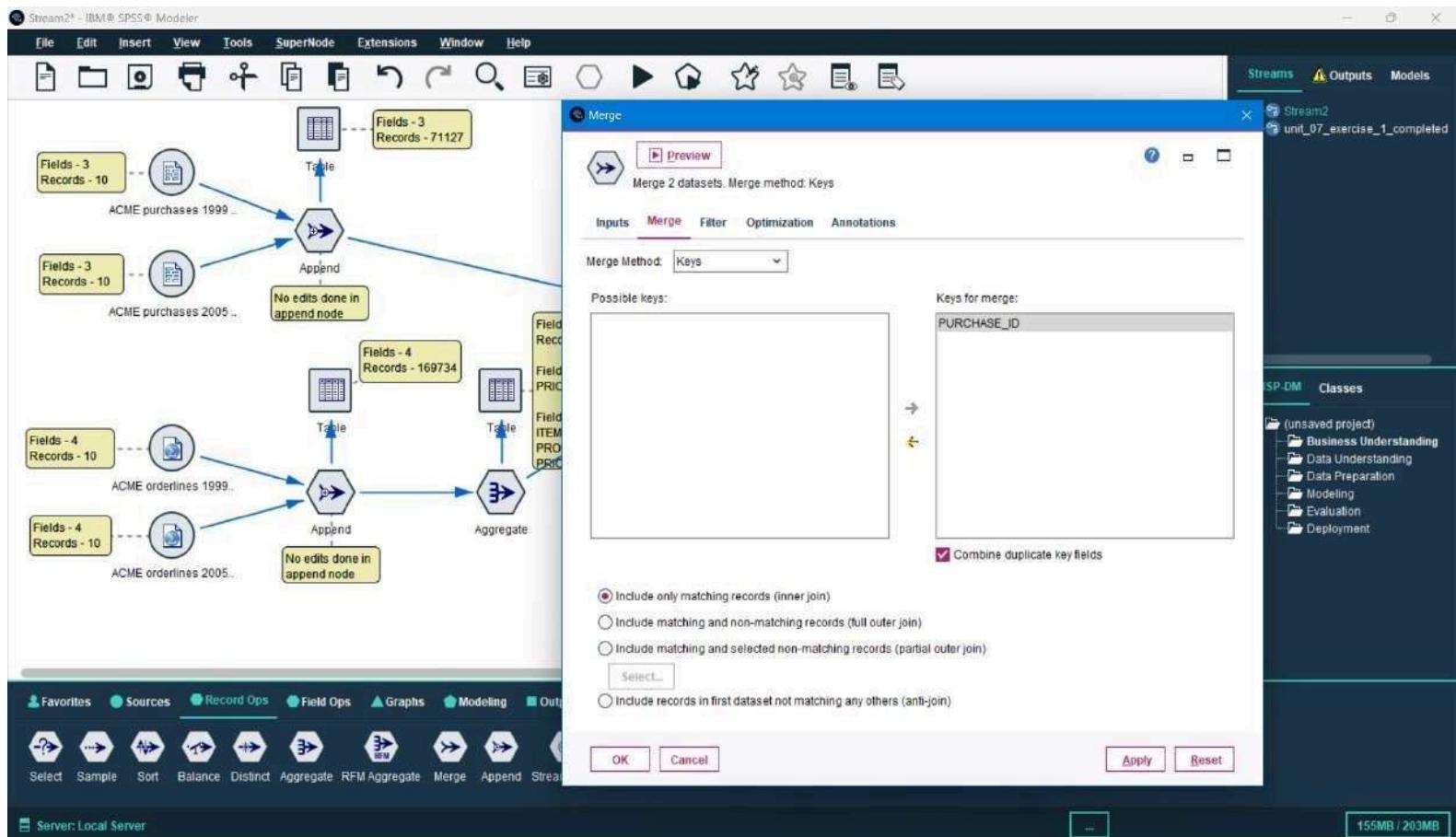
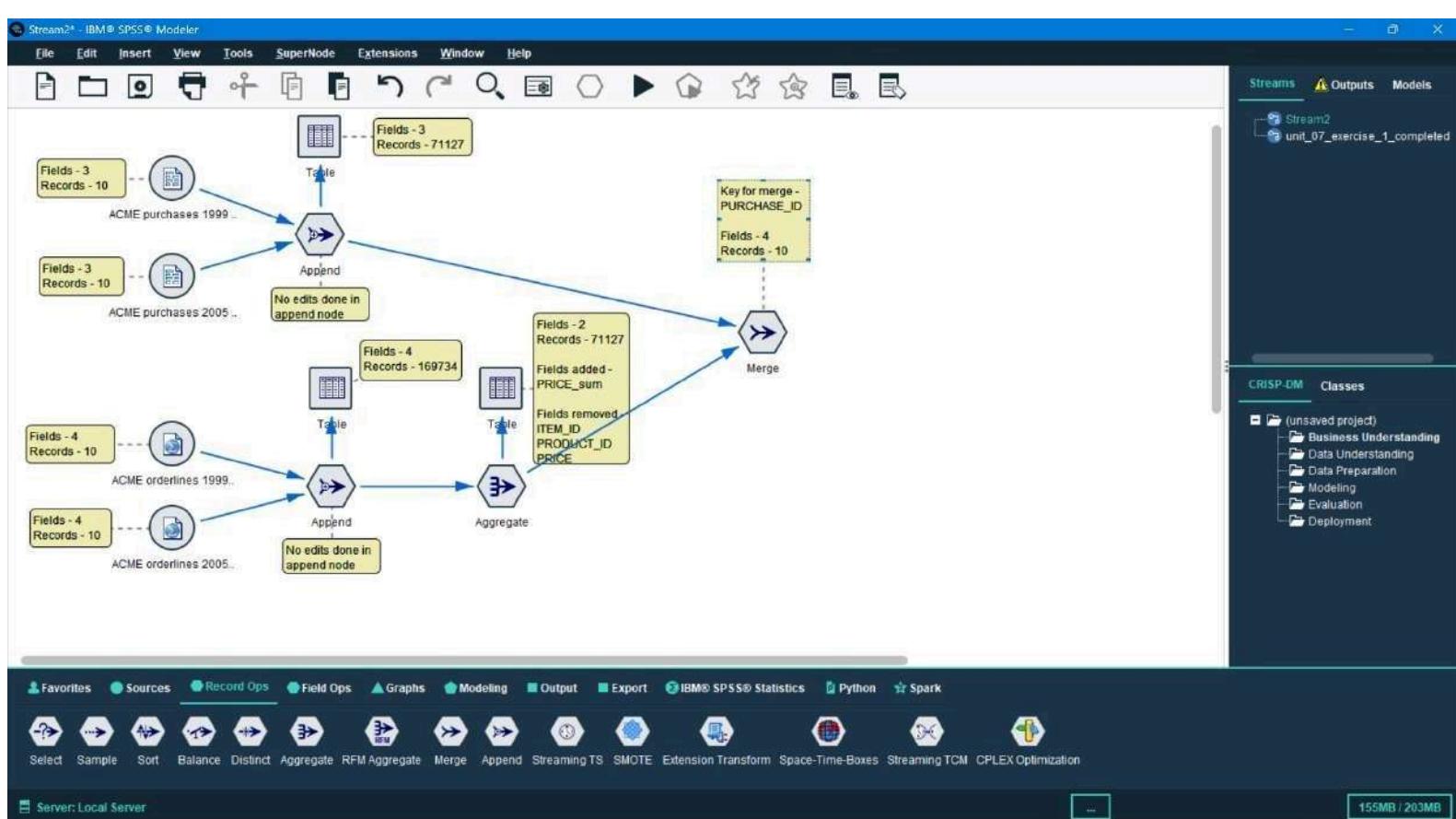
Table (2 fields, 71,127 records) #4

File Edit Generate

Table Annotations

PURCHASE_ID	PRICE_SUM
1	5723
2	5724
3	217.770
4	5727
5	5729
6	5730
7	5731
8	5732
9	5734
10	5738
11	5737
12	5738
13	5739
14	5740
15	5741
16	5743
17	5744
18	5745
19	5746
20	5747
21	5748
22	5749
23	5751
24	5753
25	5754
26	5755
27	5756
28	5758
29	5760
30	5761
31	5762
32	5763
33	5765
34	5766
35	5767
36	5768
37	5769
38	5770
39	5772
40	5773
41	5774
42	5775
43	5776
44	5777

OK

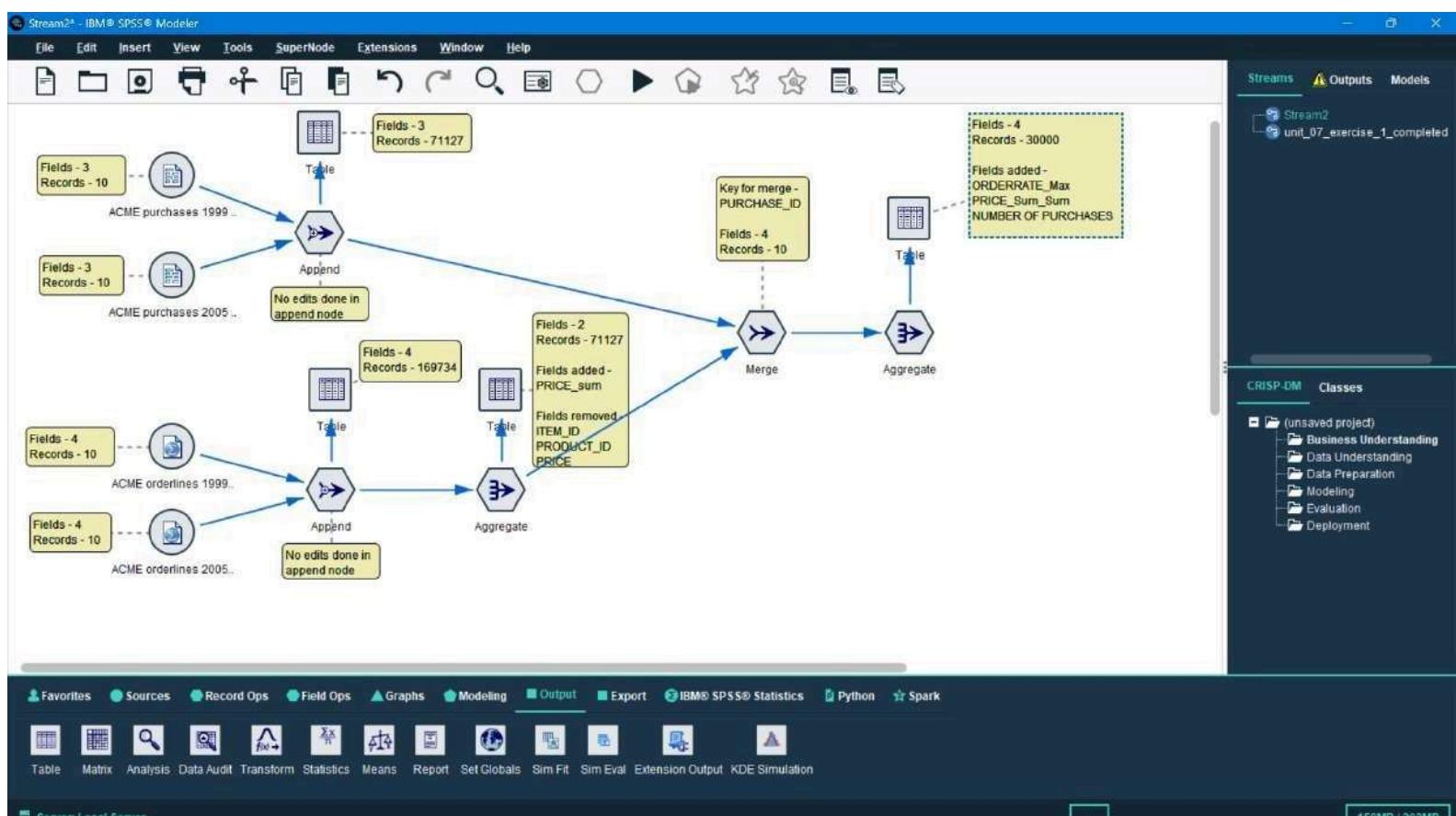


Preview from Merge Node [4 fields, 10 records] #1

File Edit Generate Table Annotations

	PURCHASE_ID	CUSTOMER_ID	ORDERDATE	PRICE_Sum
1	5723	5030	2003-06-24	948.480
2	5724	3119	2002-12-03	114.840
3	5726	2764	2002-03-13	217.770
4	5727	2190	2003-01-21	466.130
5	5729	1887	1999-07-03	985.520
6	5730	783	2002-05-26	935.350
7	5731	5584	1999-02-12	746.840
8	5732	2009	2000-10-18	940.540
9	5734	3178	2001-11-07	460.000
10	5738	5634	2003-11-18	605.410

OK



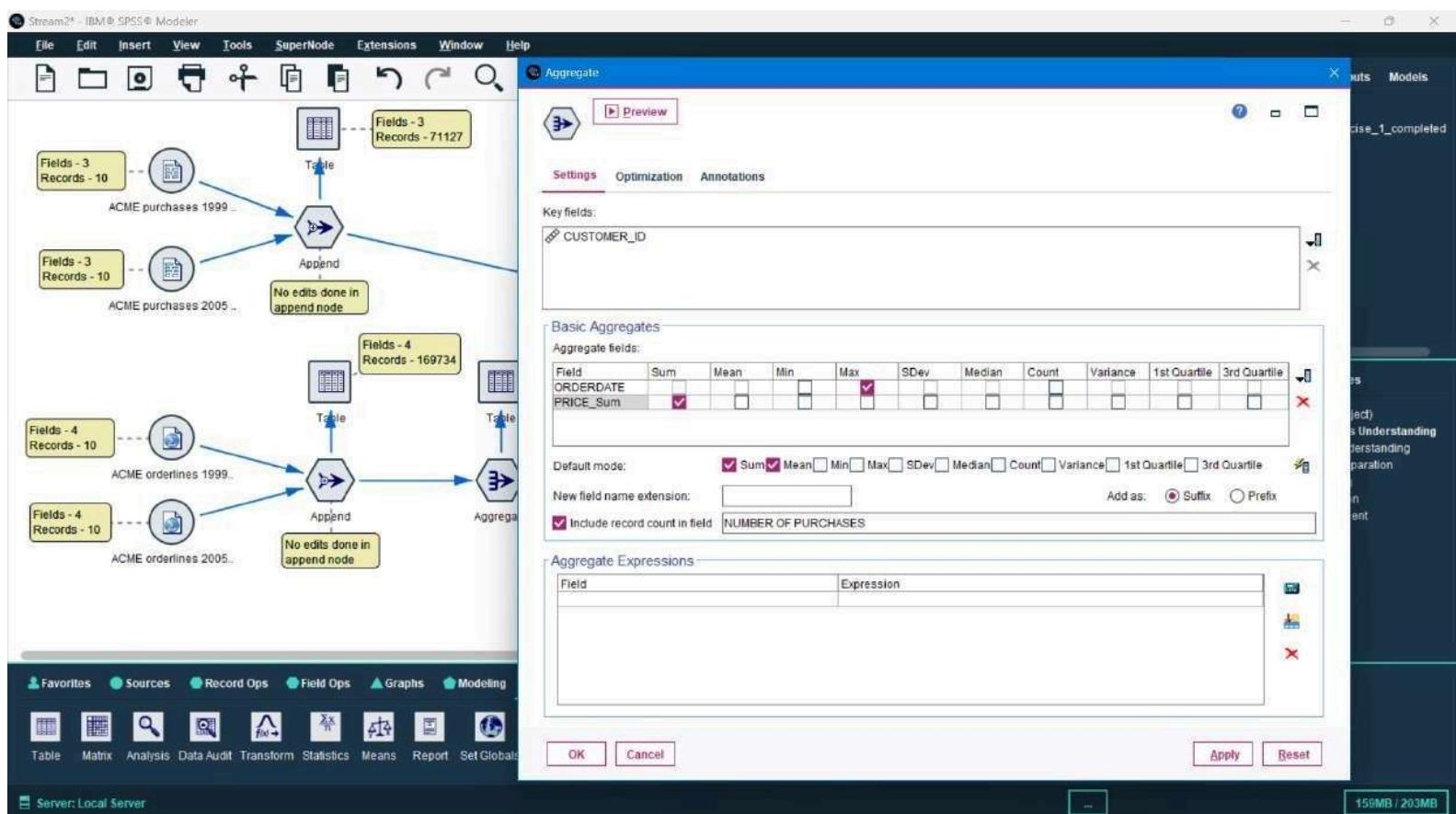
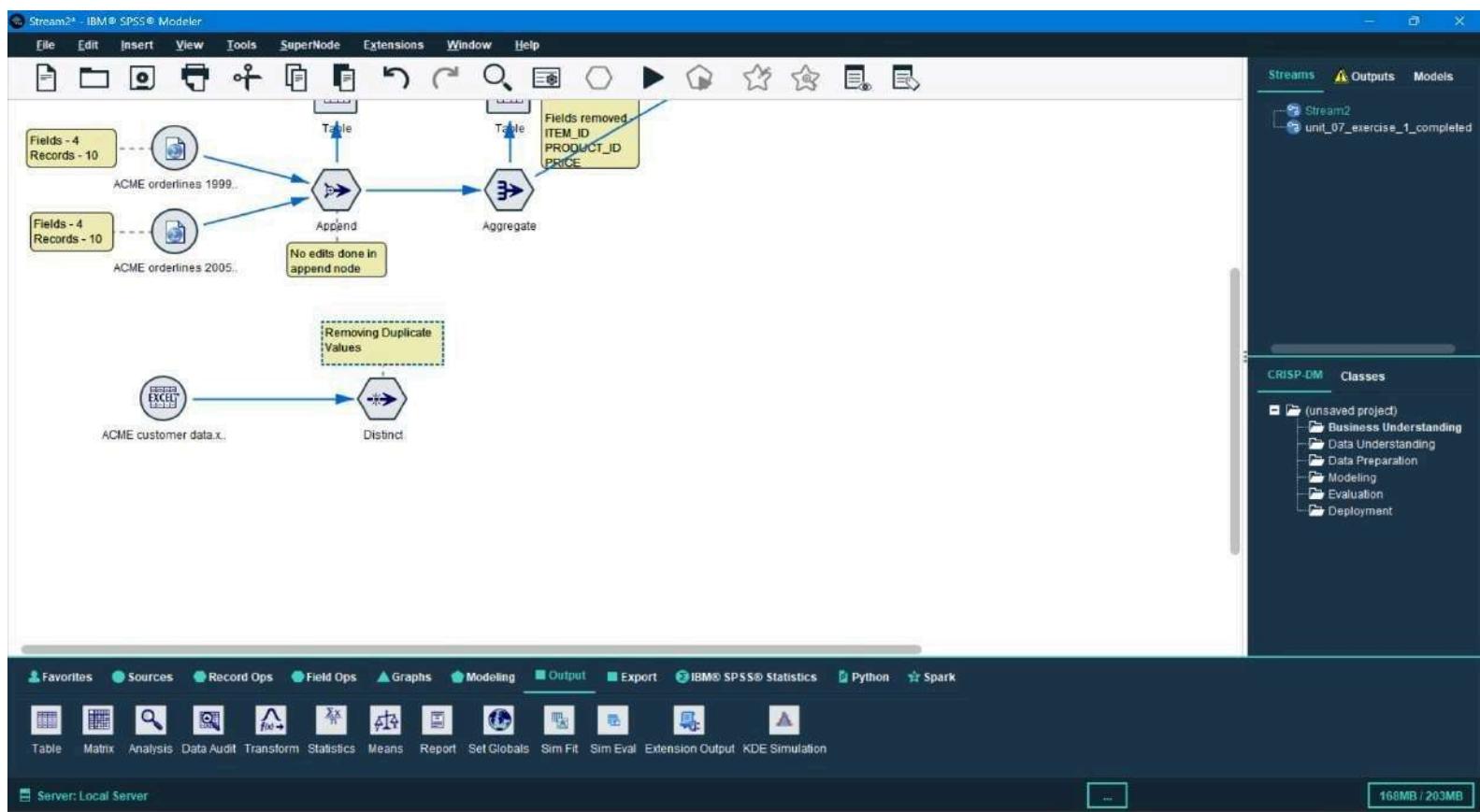


Table (4 fields, 30,000 records) #3

File Edit Generate Table Annotations

CUSTOMER_ID	ORDERDATE_Max	PRICE_Sum_Sum	NUMBER OF PURCHASES
1	5030 2003-06-24	1034.330	2
2	3119 2002-12-03	114.840	1
3	2764 2002-03-13	217.770	1
4	2190 2003-01-21	1510.940	3
5	1887 2001-08-24	1283.120	2
6	783 2002-05-26	935.350	1
7	5584 2001-09-29	763.190	2
8	2009 2004-10-27	999.410	2
9	3178 2001-11-07	1281.830	3
10	5834 2003-11-18	605.410	1
11	2874 2003-02-03	1199.570	1
12	3048 2003-02-15	922.130	2
13	2081 2000-08-13	432.150	2
14	3147 2002-10-10	1400.740	2
15	3253 2004-09-26	985.120	2
16	2454 2004-02-10	1686.300	4
17	5005 2000-01-04	75.520	1
18	5410 1999-07-04	60.560	1
19	4310 2003-09-23	501.280	1
20	3684 2002-05-14	1590.760	3
21	3386 2000-05-13	1279.270	1
22	3829 2003-06-07	288.630	1
23	5545 2003-02-09	744.840	1
24	4892 2003-12-26	2824.190	3
25	3399 2003-10-28	1292.440	3
26	5446 2001-05-01	845.100	2
27	5018 2001-01-18	1085.040	3
28	2107 2003-11-02	141.570	1
29	1175 2001-10-06	708.630	2
30	1582 2002-09-17	1425.910	1
31	5368 2001-12-05	300.300	1
32	5486 2000-05-27	312.870	2
33	3676 2003-06-11	1498.860	2
34	3963 2002-04-01	1987.160	3
35	1231 2003-05-23	360.610	3
36	3984 2002-03-30	984.380	1
37	4477 2000-11-28	840.800	1
38	4205 1999-02-25	612.820	1
39	3949 2000-05-30	350.000	1
40	861 2003-07-15	1576.670	3
41	3435 2003-09-22	1013.370	2
42	1844 2003-02-25	737.750	2
43	4903 2001-09-16	228.830	1
44	3033 2001-12-01	121.900	1



Preview from ACME customer data.xlsx Node (5 fields, 10 records) #2

	CUSTOMER_ID	GENDER	CREDITLIMIT	ZODIAC	E-MAIL ADDRESS	ZIP
1	723.000	F	9026.000	1.000	name7502@tnet.fr	1818B0
2	724.000	M	5223.000	3.000	name25485@wwwmail.org	1132DG
3	725.000	F	668.000	8.000	name15543@wwwmail.de	1803YT
4	726.000	M	8302.000	5.000	name28335@bigzag.be	1205WR
5	727.000	F	997.000	8.000	name5354@tnet.jp	1711ON
6	728.000	M	1965.000	3.000	name20637@wwwmail.es	1065FG
7	729.000	F	6898.000	5.000	name20636@wwwmail.es	1254MR
8	730.000	F	6488.000	11.000	name10414@tnet.inc	1723DG
9	730.000	F	6488.000	11.000	name10414@tnet.inc	1723DG
10	730.000	F	6488.000	11.000	name10414@tnet.inc	1723DG

Stream24 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Fields - 4 Records - 10 ACME orderlines 1999.. Fields removed ITEM_ID PRODUCT_ID PRICE

Fields - 4 Records - 10 ACME orderlines 2005.. Append No edits done append node

ACME customer data.x.

Distinct

Settings Composite Optimization Annotations

Mode: Include only the first record in each group

Key fields for grouping:

- CUSTOMER_ID
- GENDER
- CREDITLIMIT
- ZODIAC

Sort Order

Within groups, sort records by:

Field	Order

Default sort order: Ascending Descending

If no sort fields are selected then the order of records within each group is undefined.

OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 168MB / 203MB

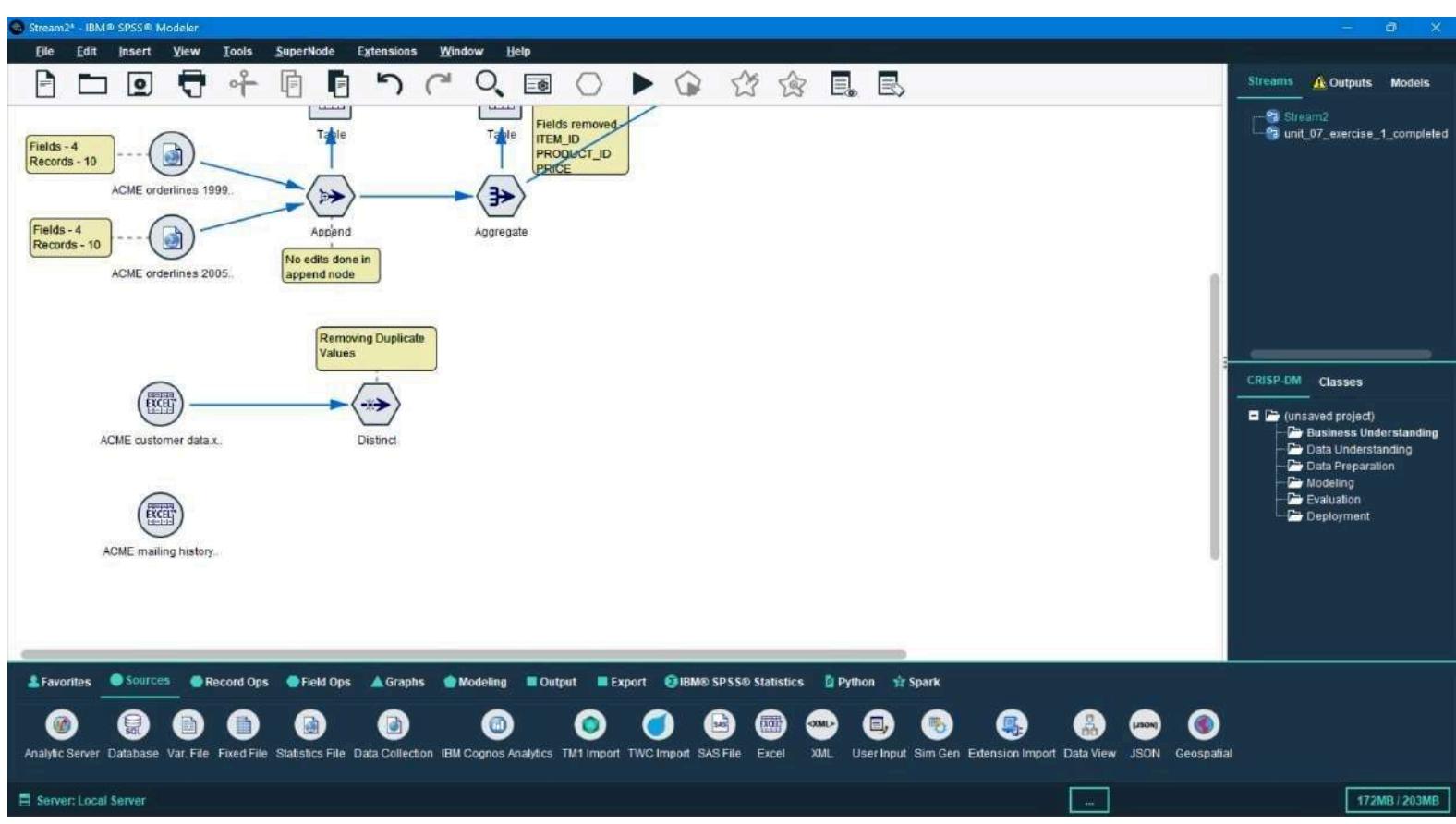
Preview from Distinct Node (6 fields, 10 records) #1

File Edit Generate

Table Annotations

	CUSTOMER_ID	GENDER	CREDITLIMIT	ZODIAC	E-MAIL ADDRESS	ZIP
1	723.000	F	9026.000	1.000	name7502@tnet.fr	1818B0
2	724.000	M	5223.000	3.000	name25485@wwwmail.org	1132DG
3	725.000	F	668.000	8.000	name15543@wwwmail.de	1803YT
4	726.000	M	8302.000	6.000	name28335@zigzag.be	1205WR
5	727.000	F	997.000	8.000	name5354@tnet.jp	1711ON
6	728.000	M	1955.000	3.000	name20637@wwwmail.es	1055FG
7	729.000	F	6898.000	5.000	name20636@wwwmail.es	1254MR
8	730.000	F	6488.000	11.000	name10414@tnet.inc	1723DG
9	731.000	F	5006.000	2.000	name23372@wwwmail.inc	1713AQ
10	732.000	F	929.000	5.000	name20635@wwwmail.es	1264EC

OK



Preview from ACME mailing history.xlsx Node (2 fields, 10 records) #2

File Edit Generate Table Annotations

	customer_id mailing
1	723.000 XL original orange baseball cap
2	723.000 XS original red soccer shirt
3	724.000 XL original orange baseball cap
4	725.000 XL original orange baseball cap
5	726.000 XL original orange baseball cap
6	727.000 Standard tennisracket
7	727.000 XL original orange baseball cap
8	727.000 XS original red soccer shirt
9	728.000 XL original orange baseball cap
10	729.000 XL original orange baseball cap

Stream2 - IBM SPSS Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

CRISP-DM Classes

Fields - 4 Records - 10 ACME orderlines 1999... Table Append Aggregate Fields removed ITEM_ID PRODUCT_ID PRICE No edits done in append node

Fields - 4 Records - 10 ACME orderlines 2005... Table

ACME customer data.x. Removing Duplicate Values Distinct

ACME mailing history. Type

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Star Rank

Auto Data Prep Type Filter Derive Filter Redclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproject Time Intervals

Server: Local Server 173MB / 203MB

Stream2 - IBM SPSS Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

CRISP-DM Classes

Fields - 4 Records - 10 ACME orderlines 1999... Table Append Aggregate Fields removed ITEM_ID PRODUCT_ID PRICE No edits done in append node

Fields - 4 Records - 10 ACME orderlines 2005... Table

ACME customer data.x. Removing Duplicate Values Distinct

ACME mailing history. Type

Type

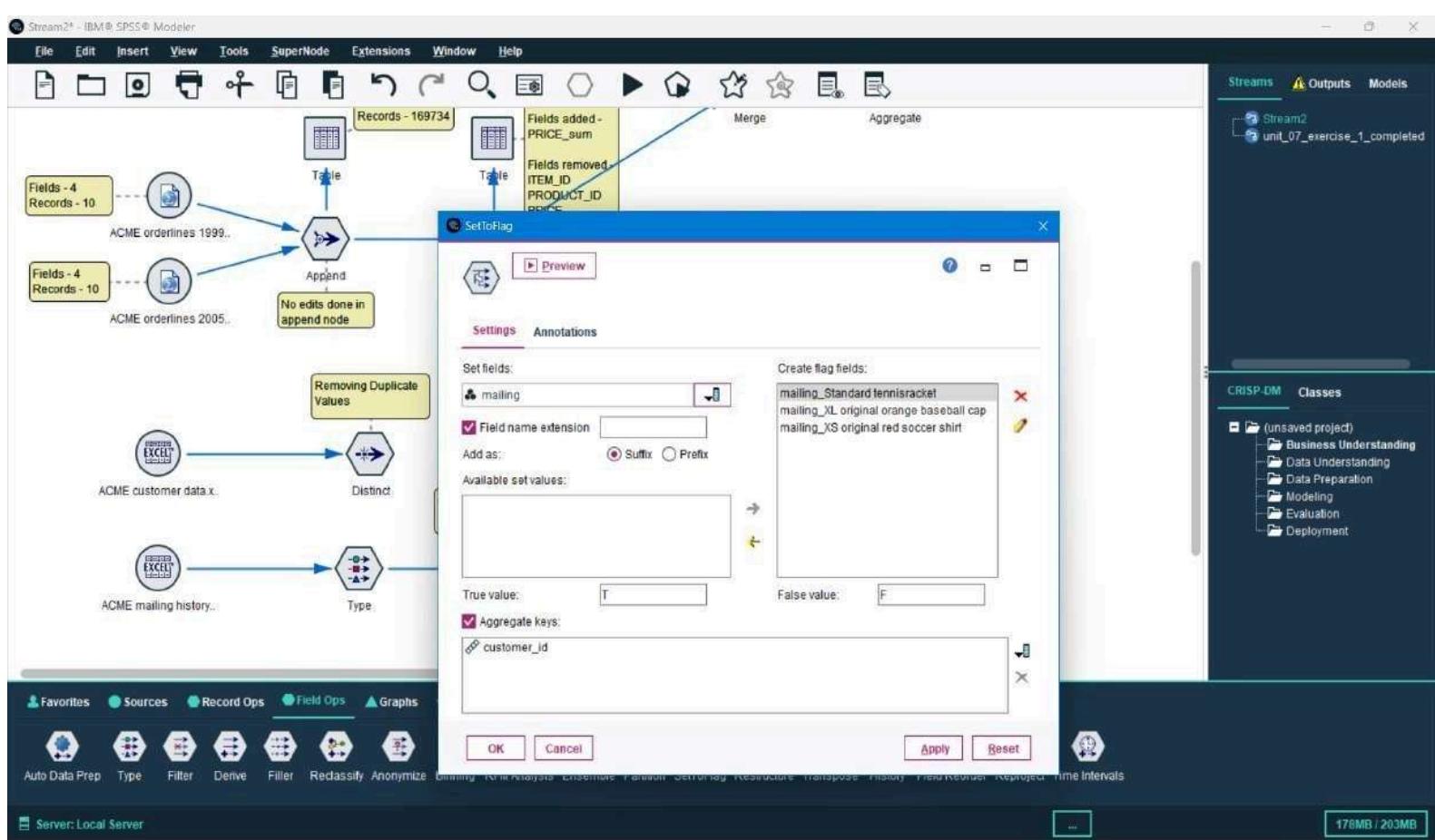
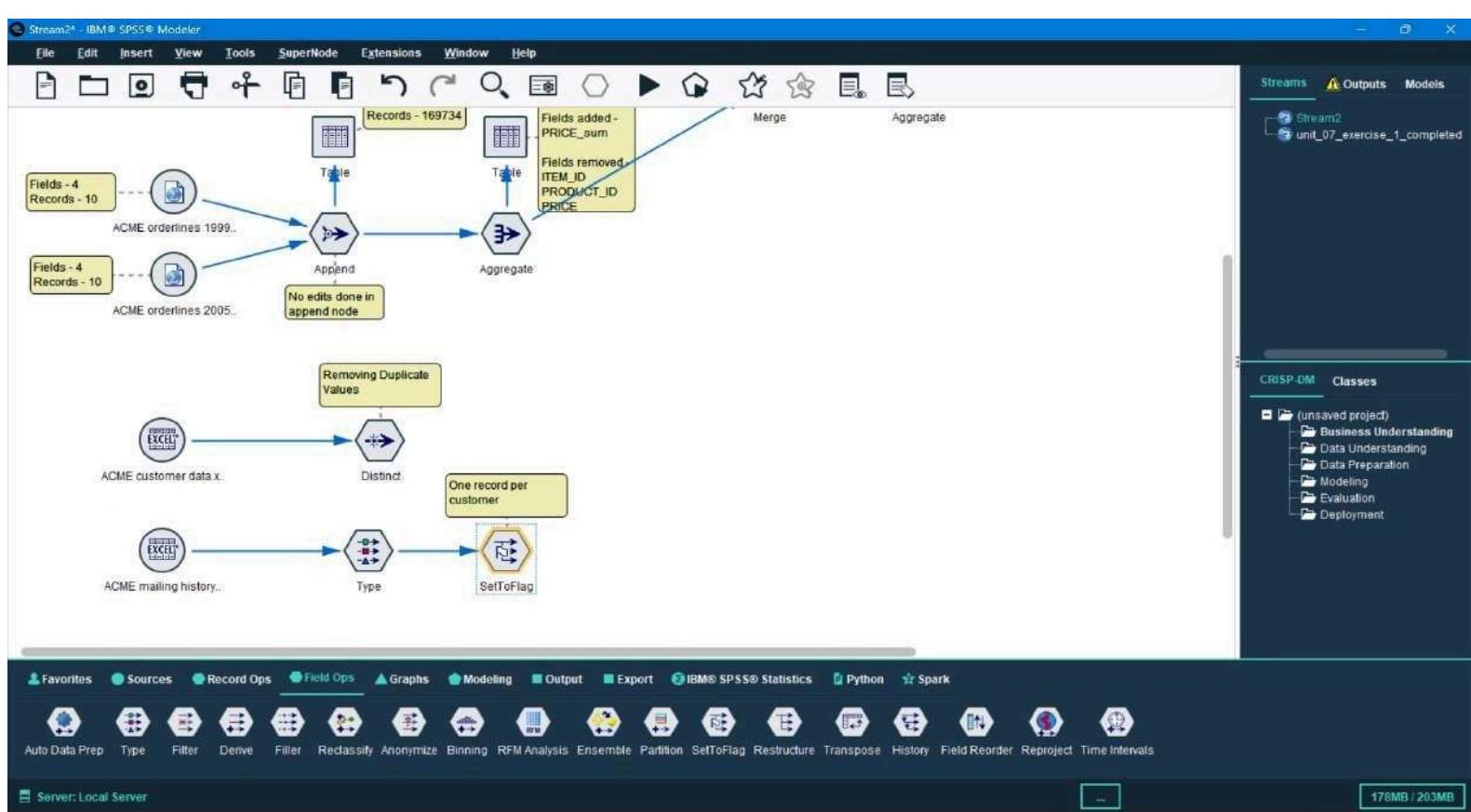
Field	Measurement	Values	Missing	Check	Role
customer_id	Continuous	<Read>	None	Input	
mailing	Nominal	<Read>	None	Input	

OK Cancel Apply Reset

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Star Rank

Auto Data Prep Type Filter Derive Filter Redclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproject Time Intervals

Server: Local Server 173MB / 203MB



Stream2 - IBM SPSS Modeler

The Stream2 interface shows a data flow starting with two Excel source nodes (ACME orderlines 1999.. and ACME orderlines 2005..) feeding into an Append node. The Append node has annotations: "No edits done in append node". The output of the Append node goes to a Merge node, which then connects to an Aggregate node. The Aggregate node has annotations: "Fields added - PRICE_sum" and "Fields removed ITEM_ID, PRODUCT_ID, PRICE". The output of the Aggregate node then flows through a "Removing Duplicate Values" node (Distinct), followed by a "One record per customer" node (Type). Finally, the data passes through a SetToFlag node and a Filter node. The Filter node has annotations: "Fields: 4 in, 0 filtered, 1 renamed, 4 out". The output of the Filter node is then processed by a Merge node and an Aggregate node.

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Auto Data Prep Type Filter Derive Filter Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproject Time Intervals

Server: Local Server 179MB / 203MB

Stream2 - IBM SPSS Modeler

The Stream2 interface shows a data flow starting with two Excel source nodes (ACME orderlines 1999.. and ACME orderlines 2005..) feeding into an Append node. The Append node has annotations: "No edits done in append node". The output of the Append node goes to a Merge node, which then connects to an Aggregate node. The Aggregate node has annotations: "Fields added - PRICE_sum" and "Fields removed ITEM_ID, PRODUCT_ID, PRICE". The output of the Aggregate node then flows through a "Removing Duplicate Values" node (Distinct), followed by a "One record per customer" node (Type). Finally, the data passes through a SetToFlag node and a Filter node. The Filter node has annotations: "Fields: 4 in, 0 filtered, 1 renamed, 4 out". The output of the Filter node is then processed by a Merge node and an Aggregate node.

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Auto Data Prep Type Filter Derive Filter Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproject Time Intervals

Server: Local Server 179MB / 203MB

Filter

Preview

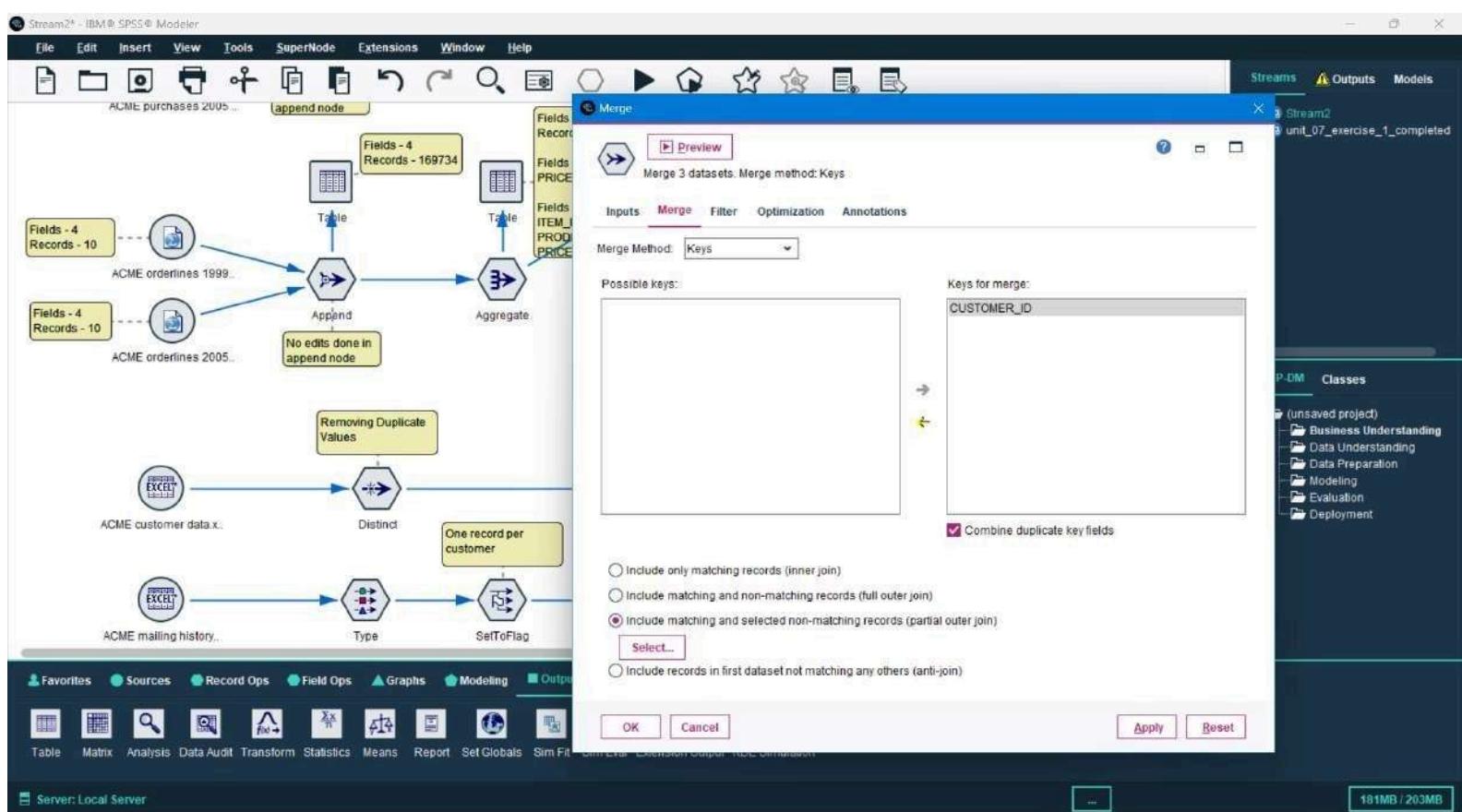
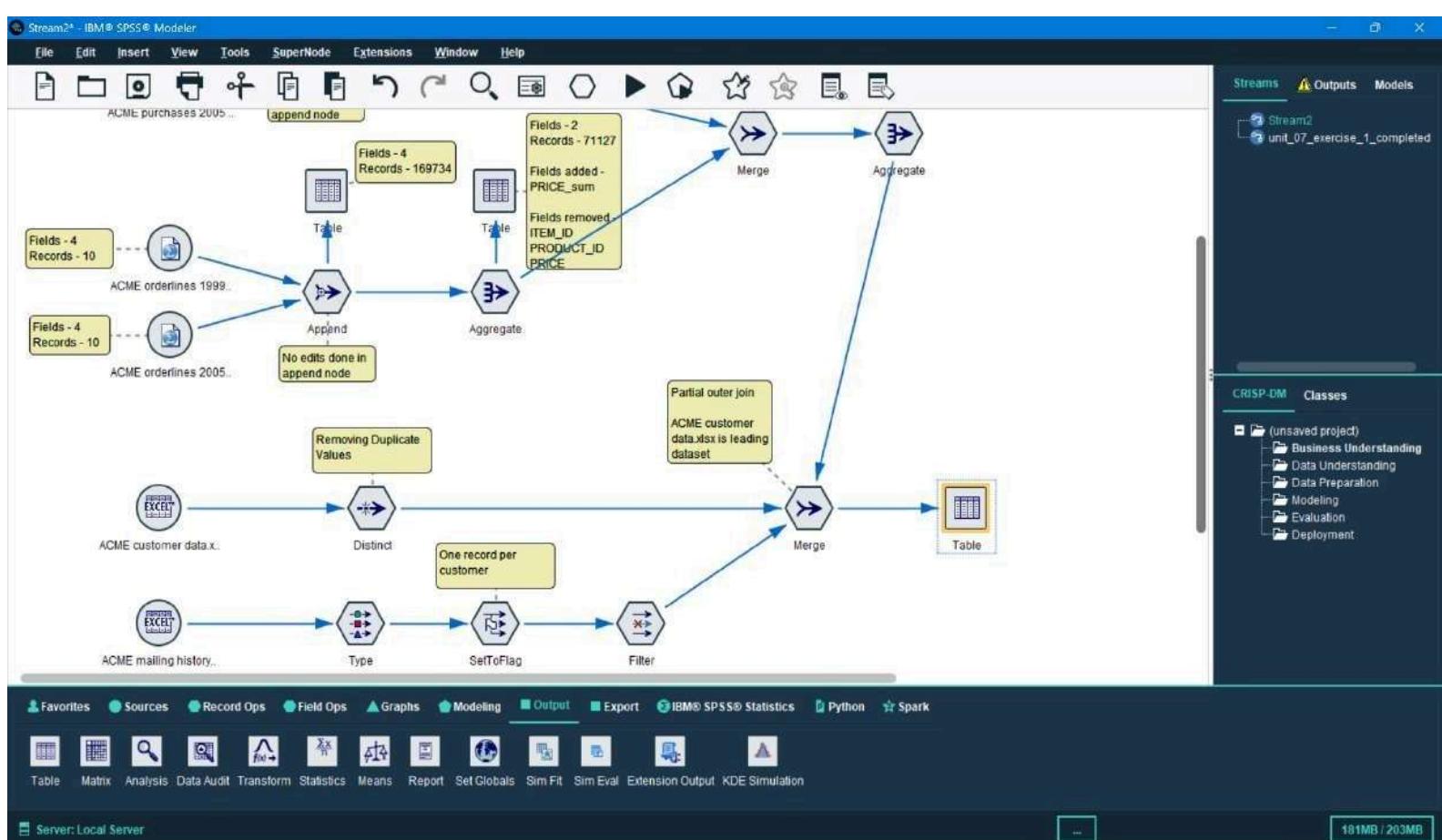
Filter Annotations

Fields: 4 in, 0 filtered, 1 renamed, 4 out

Field	Filter	Field
customer_id	CUSTOMER_ID	
mailing_Standard tennisracket	mailing_Standard tennisracket	
mailing_XL original orange ba...	mailing_XL original orange ba...	
mailing_XS original red socce...	mailing_XS original red socce...	

View current fields View unused field settings

OK Cancel Apply Reset



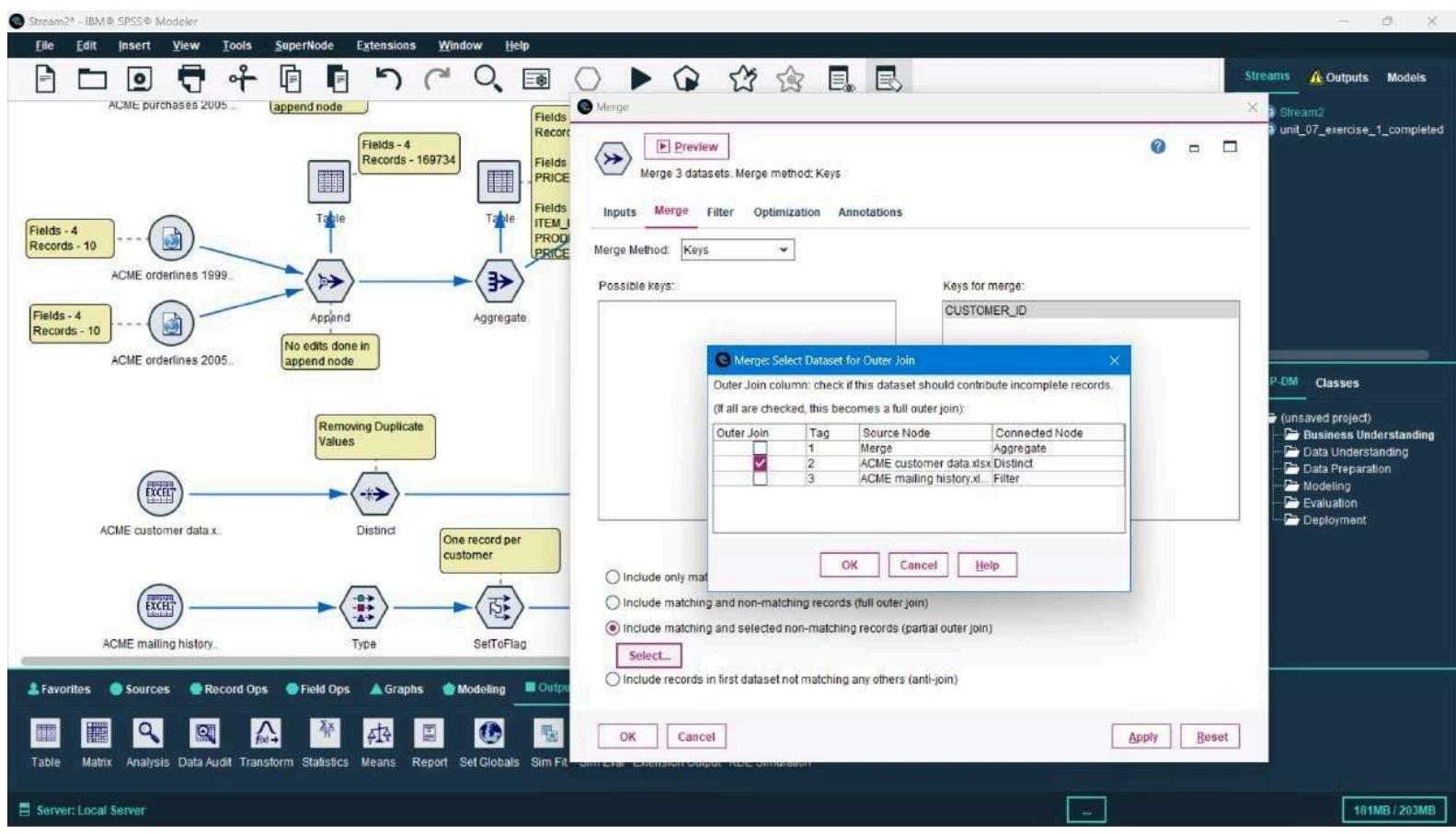


Table (12 fields, 30,000 records)

	CUSTOMER_ID	ORDERDATE_Max	PRICE_Sum_Sum	NUMBER_OF_PURCHASES	GENDER	CREDITLIMIT	ZODIAC	E-MAIL_ADDRESS	ZIP	mailing_Standard tennisracket	mailing_XL original orange baseball cap	mailing_XS original red soc
1	723,000	2009-07-02	546,731	2	F	9025,000	1,000	name25485@wwwmail.fr	1818BO	F	T	F
2	724,000	1999-07-11	306,460	1	M	5223,000	3,000	name25485@wwwmail.org	1132DG	F	T	F
3	725,000	2007-02-27	1796,016	4	F	886,000	8,000	name15543@wwwmail.de	1803YR	F	T	F
4	726,000	2010-06-14	2489,259	6	M	8302,000	6,000	name28335@zigzag.be	1205WR	F	T	F
5	727,000	1999-02-21	377,210	1	F	997,000	8,000	name53549@net.jp	1711OH	T	T	T
6	728,000	2004-05-22	624,015	1	M	1955,000	3,000	name20637@wwwmail.es	1055FG	F	T	F
7	729,000	2006-07-13	34,724	1	F	6898,000	5,000	name20636@wwwmail.es	1254M	F	T	F
8	730,000	2002-11-29	788,150	1	F	6488,000	11,000	name10414@inetinc	1723DG	T	T	F
9	731,000	2003-07-17	2413,870	3	F	5005,000	2,000	name23372@wwwmail.inc	1713AO	F	T	T
10	732,000	2003-05-10	1158,180	2	F	929,000	5,000	name20635@wwwmail.es	1284EC	F	T	T
11	733,000	2003-12-01	888,330	2	F	3303,000	8,000	name53549@net.jp	1648BT	F	T	T
12	734,000	2010-07-28	1036,026	3	F	8934,000	3,000	name17582@wwwmail.de	1285V	F	T	T
13	735,000	2009-05-26	457,036	3	F	9808,000	7,000	name6388@net.fr	1282NB	F	T	F
14	736,000	2003-10-22	1554,910	2	M	4723,000	6,000	name10409@inetinc	1799IT	F	T	T
15	737,000	2003-03-09	544,040	1	M	672,000	7,000	name13849@frat.uk	1802UO	T	T	F
16	738,000	2002-11-13	406,510	1	F	7801,000	8,000	name25473@wwwmail.es	1971NK	F	T	F
17	739,000	2003-01-28	484,000	1	M	5984,000	5,000	name13848@frat.uk	1361RL	F	T	F
18	740,000	2001-02-05	1971,720	2	F	2621,000	9,000	name23366@wwwmail.inc	1164VN	T	T	F
19	741,000	2001-10-04	1316,750	2	M	6062,000	8,000	name3188@molbe.cat	1767YN	F	T	F
20	742,000	2000-01-22	386,320	1	M	9797,000	4,000	name15086@omejor.es	1681HP	F	T	T
21	743,000	2003-09-16	1454,310	2	F	2987,000	6,000	name29056@zigzag.cat	1287P	F	T	T
22	744,000	2004-08-09	711,319	1	M	128,000	9,000	name6323@zigzag.be	1023AR	F	T	F
23	745,000	2001-10-20	1199,380	2	F	182,000	11,000	name20630@wwwmail.es	1737ZI	F	T	F
24	746,000	2008-01-28	165,638	2	M	2301,000	9,000	name13847@frat.uk	1625TZ	T	T	F
25	747,000	2008-08-15	207,407	1	F	2704,000	8,000	name10401@inetinc	1684YE	T	T	F
26	748,000	2003-08-27	785,060	2	M	4033,000	7,000	name13846@frat.uk	1508DU	F	T	F
27	749,000	2009-02-08	128,067	2	M	1818,000	9,000	name17579@wwwmail.de	1143C	F	T	F
28	750,000	2009-02-23	1365,367	3	F	9044,000	6,000	name25894@wwwmail.uk	1228IK	F	T	F
29	751,000	1999-10-24	1706,760	2	F	1975,000	12,000	name6366@inet.fr	1567CX	F	T	F
30	752,000	2002-12-13	1757,480	2	F	3619,000	9,000	name1509@omejor.es	1315FI	F	T	F
31	753,000	2002-01-18	454,800	1	M	3517,000	8,000	name29727@zigzag.com	1753OI	F	T	T
32	754,000	2010-04-29	405,287	2	M	701,000	8,000	name29856@zigzag.es	1229DC	F	T	F
33	755,000	2004-05-14	573,874	1	M	1878,000	3,000	name17578@wwwmail.de	1038LJ	F	T	F
34	756,000	2004-11-15	3215,550	3	M	4023,000	10,000	name29046@zigzag.cat	1002W	F	T	F
35	757,000	2000-05-01	959,500	1	M	4946,000	10,000	name10389@inetinc	1979KZ	T	T	F
36	758,000	2010-10-30	793,894	3	F	9946,000	8,000	name63546@net.jp	1784IE	F	T	F
37	759,000	2003-03-04	1183,400	2	F	1522,000	6,000	name3172@molbe.cat	1183FC	F	T	F
38	760,000	2001-11-19	21,560	1	F	7456,000	2,000	name426hmail.de	1492WX	F	T	F
39	761,000	2002-11-24	1195,590	3	F	1340,000	10,000	name5357@inet.jp	1569Y	F	T	F
40	762,000	2002-12-10	286,970	1	F	7136,000	3,000	name2556@hmail.in	1558OG	F	T	F
41	763,000	2011-06-30	1687,274	1	F	1723,000	4,000	name25457@wwwmail.org	1588SB	F	T	F
42	764,000	1999-09-07	241,700	1	F	3799,000	7,000	name1516@omejor.es	1972BQ	F	T	F
43	765,000	2002-09-21	811,531	2	M	1326,000	2,000	nameF349@inet.in	1118QV	F	T	F

OK

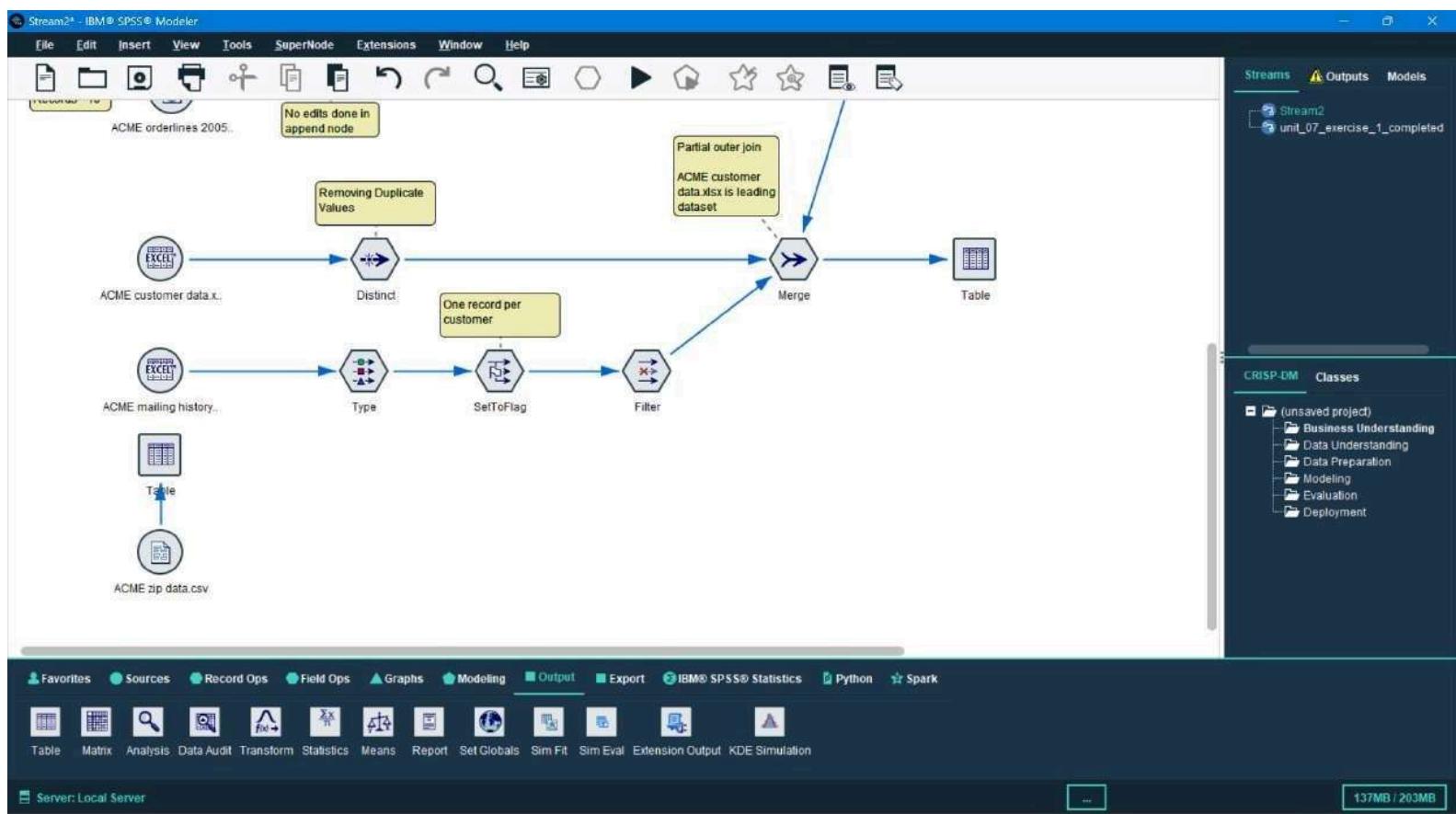


Table (4-fields, 4,083 records) #3

File Edit Generate

	ZIP	pct_working	pct_house_owners	pct_with_car
1	1001EZ	76	60	87
2	1001FA	76	60	88
3	1001GB	70	48	88
4	1001PJ	76	59	90
5	1001PR	86	69	91
6	1002AJ	88	72	92
7	1002EH	76	59	92
8	1002GZ	79	67	91
9	1002JC	76	57	87
10	1002JX	90	64	92
11	1002MP	83	67	91
12	1002WV	83	70	92
13	1002WX	80	67	92
14	1002YB	73	57	93
15	1003IX	71	56	89
16	1003OH	80	62	90
17	1003QN	88	74	90
18	1003TK	91	72	92
19	1003UB	67	49	93
20	1004CR	80	60	88
21	1004IB	81	65	89
22	1004PO	84	70	94
23	1004KV	85	70	88
24	1005GB	81	64	87
25	1005GR	77	62	88
26	1005SZ	70	55	95
27	1006FL	84	71	94
28	1006KL	70	54	96
29	1006VP	75	63	90
30	1006WP	75	61	92
31	1007AE	74	58	85
32	1007HU	76	63	86
33	1007JU	82	65	86
34	1007TN	83	70	90
35	1007YV	79	64	90
36	1008CD	75	60	89
37	1008DS	82	66	90
38	1008GT	71	61	92
39	1008IP	78	65	90
40	1008KU	82	66	96
41	1008LJ	82	66	91
42	1008NV	72	57	91
43	1008YE	84	69	89
44	1008YS	81	64	90

OK

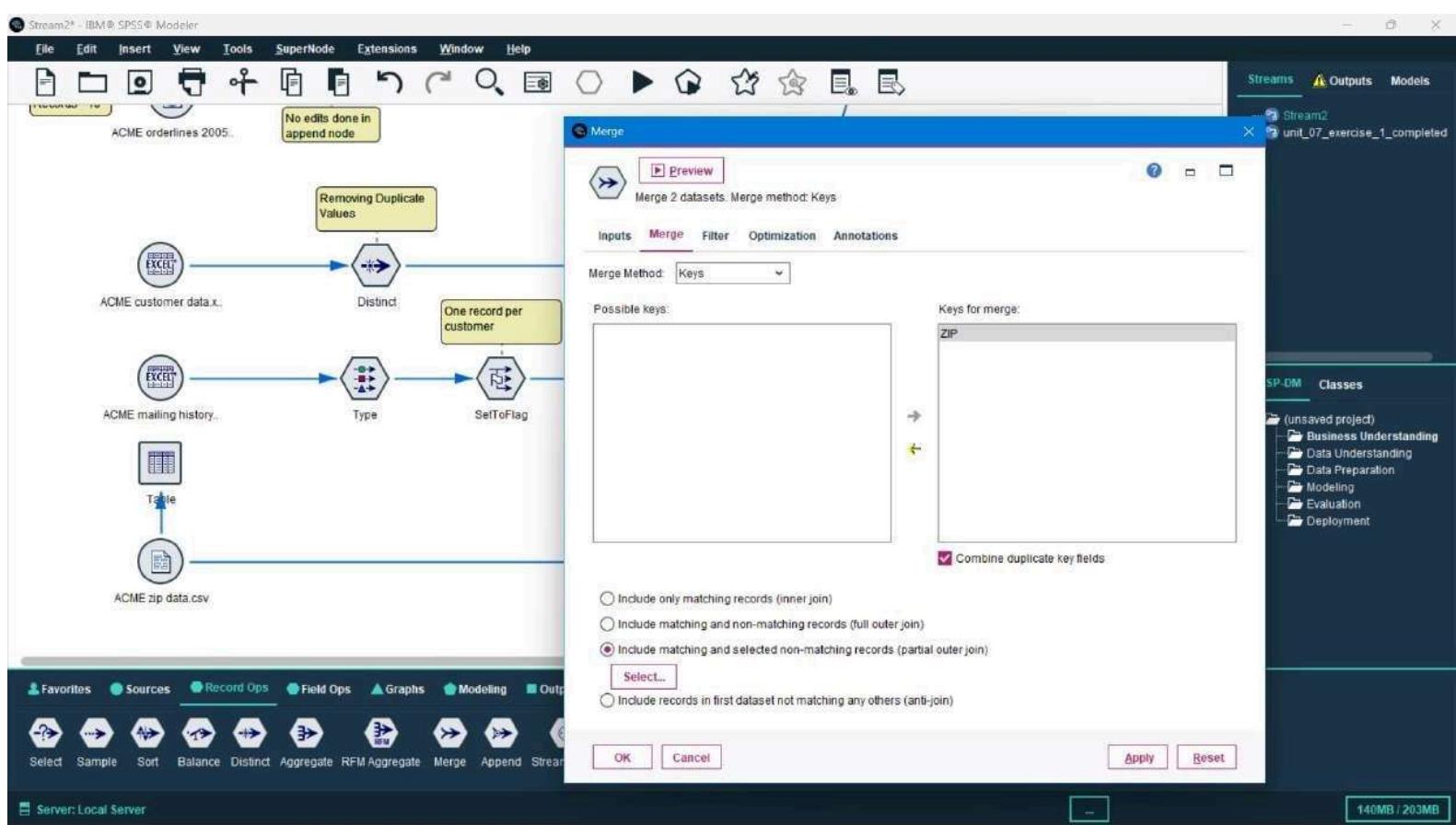
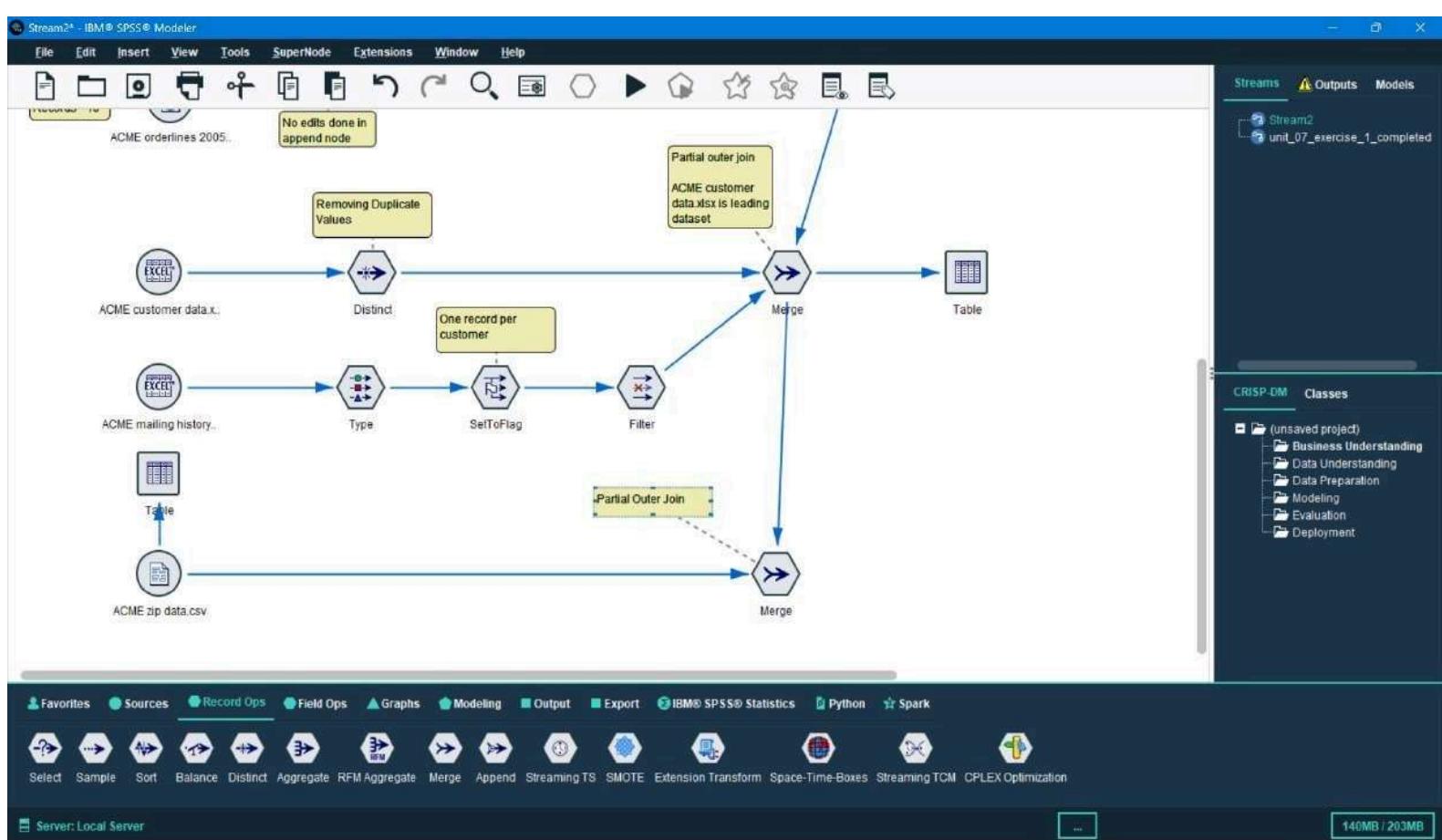
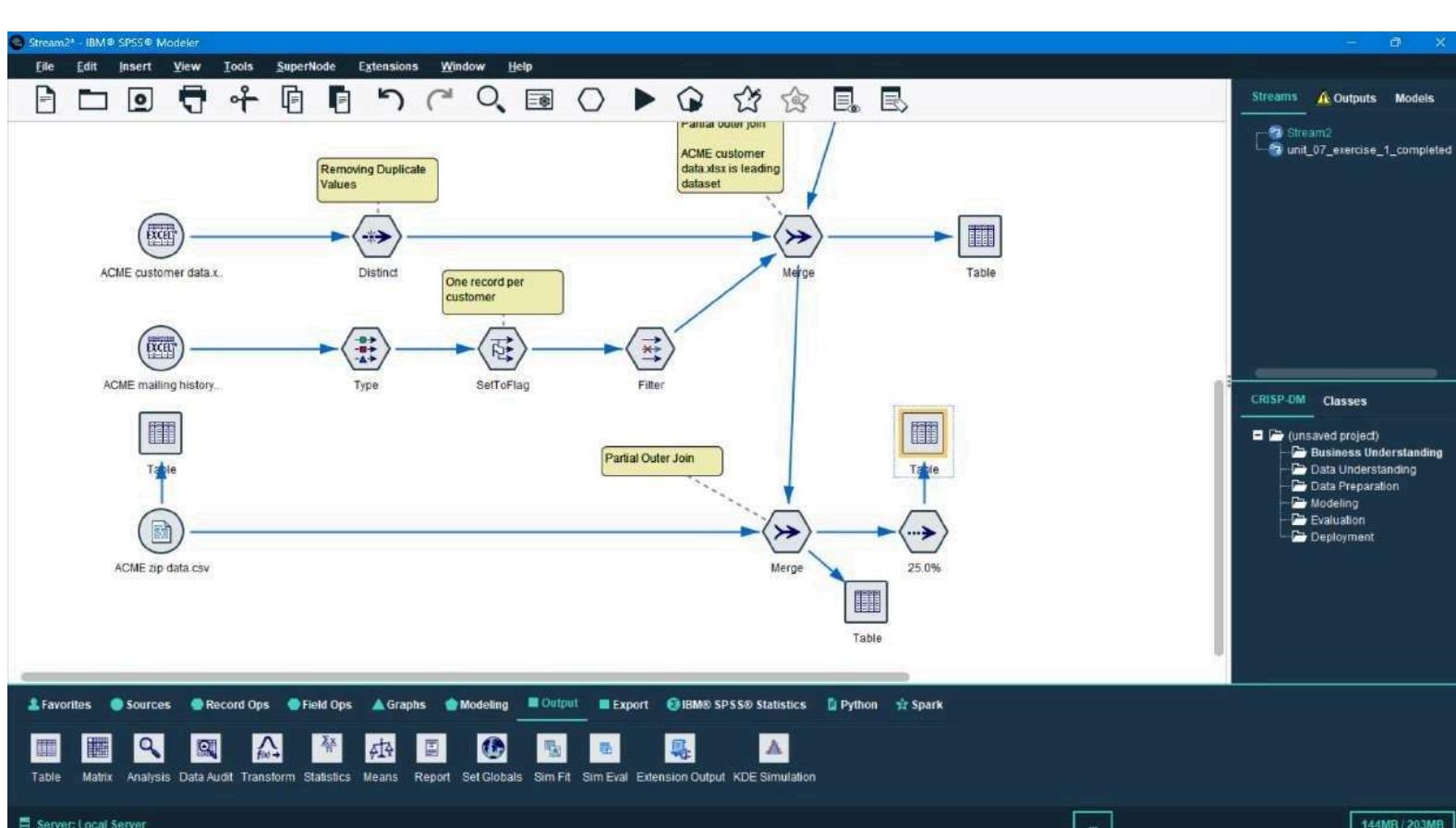


Table (15 fields, 29,685 records)																
Table		Annotations														
ZIP	pct_working	pct_house_owners	pct_with_car	CUSTOMER_ID	ORDERDATE_Max	PRICE_Sum_Sum	NUMBER_OF_PURCHASES	GENDER	CREDITLIMIT	ZODIAC	E-MAIL_ADDRESS	mailing_Standard	tennisracket	mailing_XL	original or	
1	1001EZ	76	60	87	9016.000	2009-01-15	552.667	2F	3692.000	1.000.name2048@lomejor.es	\$null\$	\$null\$				
2	1001EZ	76	60	87	14016.000	2007-07-27	632.249	1F	3692.000	4.000.name1111wwwmail.es	F					
3	1001EZ	76	60	87	4015.000	2009-03-30	1273.626	6F	3692.000	2.000.name2515@wwwmail.org	F	T				
4	1001EZ	76	60	87	19016.000	2010-12-05	904.070	4F	3692.000	3.000.names8954@netinc	\$null\$	\$null\$				
5	1001EZ	76	60	87	24016.000	2009-01-22	94.209	1F	3692.000	9.000.name6707@net.tr	\$null\$	\$null\$				
6	1001EZ	76	60	87	29016.000	2010-03-19	1614.197	5F	3692.000	8.000.name8113@netfr	F	F				
7	1001FA	76	60	88	6952.000	2009-08-20	1318.584	3F	6747.000	4.000.name2132@wwwmail.inc	\$null\$	\$null\$				
8	1001FA	76	60	88	1952.000	2009-05-02	244.742	2F	6747.000	9.000.name25794@wwwmail.uk	F	T				
9	1001FA	76	60	88	16952.000	2007-12-03	820.185	2F	6747.000	10.000.name25364@wwwmail.org	\$null\$	\$null\$				
10	1001FA	76	60	88	11952.000	2008-09-28	430.783	2F	6747.000	1.000.name8125@netfr	\$null\$	\$null\$				
11	1001FA	76	60	88	21952.000	2010-06-27	171.112	1F	6747.000	11.000.name14306@net.uk	T	F				
12	1001FA	76	60	88	26952.000	2008-04-20	2611.882	6F	6747.000	5.000.name21012@wwwmail.inc	\$null\$	\$null\$				
13	1001GB	70	48	88	13110.000	2008-01-29	795.883	2M	706.000	3.000.name9056@netinc	\$null\$	\$null\$				
14	1001GB	70	48	88	8110.000	2010-12-06	1472.307	6M	706.000	10.000.name12835@net.uk	\$null\$	\$null\$				
15	1001GB	70	48	88	18110.000	2009-05-23	755.751	2M	706.000	6.000.name27443@zlgtag.be	\$null\$	\$null\$				
16	1001GB	70	48	88	28110.000	2004-08-02	2138.859	2M	706.000	6.000.name13572@net.uk	F	T				
17	1001GB	70	48	88	28110.000	2010-06-17	274.572	2M	706.000	3.000.name7594@netfr	F	F				
18	1001GB	70	48	88	23110.000	2006-10-15	265.197	2M	706.000	5.000.name10720@net.inc	\$null\$	\$null\$				
19	1001PJ	76	59	90	5296.000	2002-03-09	1195.490	2F	2831.000	8.000.name26815@wwwmail.uk	F	T				
20	1001PJ	76	59	90	15259.000	2008-05-01	1504.776	3F	2631.000	4.000.name16931@wwwmail.de	\$null\$	\$null\$				
21	1001PJ	76	59	90	10259.000	2005-07-11	117.106	1F	2631.000	7.000.name3803@moble.cat	T	F				
22	1001PJ	76	59	90	30259.000	2010-12-07	1288.636	3F	2631.000	1.000.name14640@wwwmail.de	\$null\$	\$null\$				
23	1001PJ	76	59	90	20259.000	2009-08-08	1059.022	3F	2631.000	9.000.name4699@net.jp	\$null\$	\$null\$				
24	1001PJ	76	59	90	25259.000	2010-02-25	1828.660	5F	2631.000	6.000.name5560@net.jp	\$null\$	\$null\$				
25	1001PR	86	69	91	8293.000	2008-01-25	901.709	5M	8171.000	1.000.name18983@wwwmail.es	T	T				
26	1001PR	86	69	91	3092.000	2000-04-27	414.060	1M	8171.000	6.000.name26964@wwwmail.uk	F	T				
27	1001PR	86	69	91	18293.000	2008-04-07	539.568	2M	8171.000	3.000.name10516@net.inc	F	T				
28	1001PR	86	69	91	13293.000	2010-07-15	1656.827	6M	8171.000	9.000.name8830@netinc	F	T				
29	1001PR	86	69	91	28293.000	2007-01-10	679.607	2M	8171.000	3.000.name10035@netinc	T	T				
30	1001PR	86	69	91	23293.000	2007-01-25	1051.958	2M	8171.000	4.000.name15703@wwwmail.de	T	T				
31	1002AI	88	72	92	5597.000	1999-06-27	1178.480	2M	2198.000	3.000.name13167@net.uk	F	T				
32	1002AI	88	72	92	10560.000	2009-03-13	508.015	3M	2198.000	11.000.name22614@wwwmail.inc	T	F				
33	1002AI	88	72	92	15560.000	2009-06-13	945.739	6M	2198.000	10.000.name5601@net.jp	T	F				
34	1002AI	88	72	92	20560.000	2009-03-18	337.742	1M	2198.000	4.000.name4641@net.jp	\$null\$	\$null\$				
35	1002AI	88	72	92	25560.000	2010-03-25	1346.861	3M	2198.000	3.000.name5504@net.jp	\$null\$	\$null\$				
36	1002AI	88	72	92	30560.000	2010-08-22	1386.061	3M	2198.000	9.000.name3915@net.jp	\$null\$	\$null\$				
37	1002EH	76	59	92	11891.000	2009-07-18	549.209	2M	8927.000	4.000.name17252@wwwmail.de	T	F				
38	1002EH	76	59	92	6891.000	2008-05-18	43.413	1M	8927.000	10.000.name12993@net.uk	\$null\$	\$null\$				
39	1002EH	76	59	92	16891.000	2010-10-25	2756.640	5M	8927.000	3.000.name17381@wwwmail.de	T	F				
40	1002EH	76	59	92	1690.000	2002-01-21	847.100	2M	8927.000	11.000.name28295@zlgtag.be	F	T				
41	1002EH	76	59	92	21891.000	2009-01-02	580.442	2M	8927.000	11.000.name8832@netinc	\$null\$	\$null\$				
42	1002EH	76	59	92	26891.000	2008-08-30	828.505	3M	8927.000	5.000.name13562@net.uk	\$null\$	\$null\$				
43	1002GZ	79	67	91	9410.000	2001-11-07	120.150	1F	8464.000	3.000.name13424@net.uk	F	T				



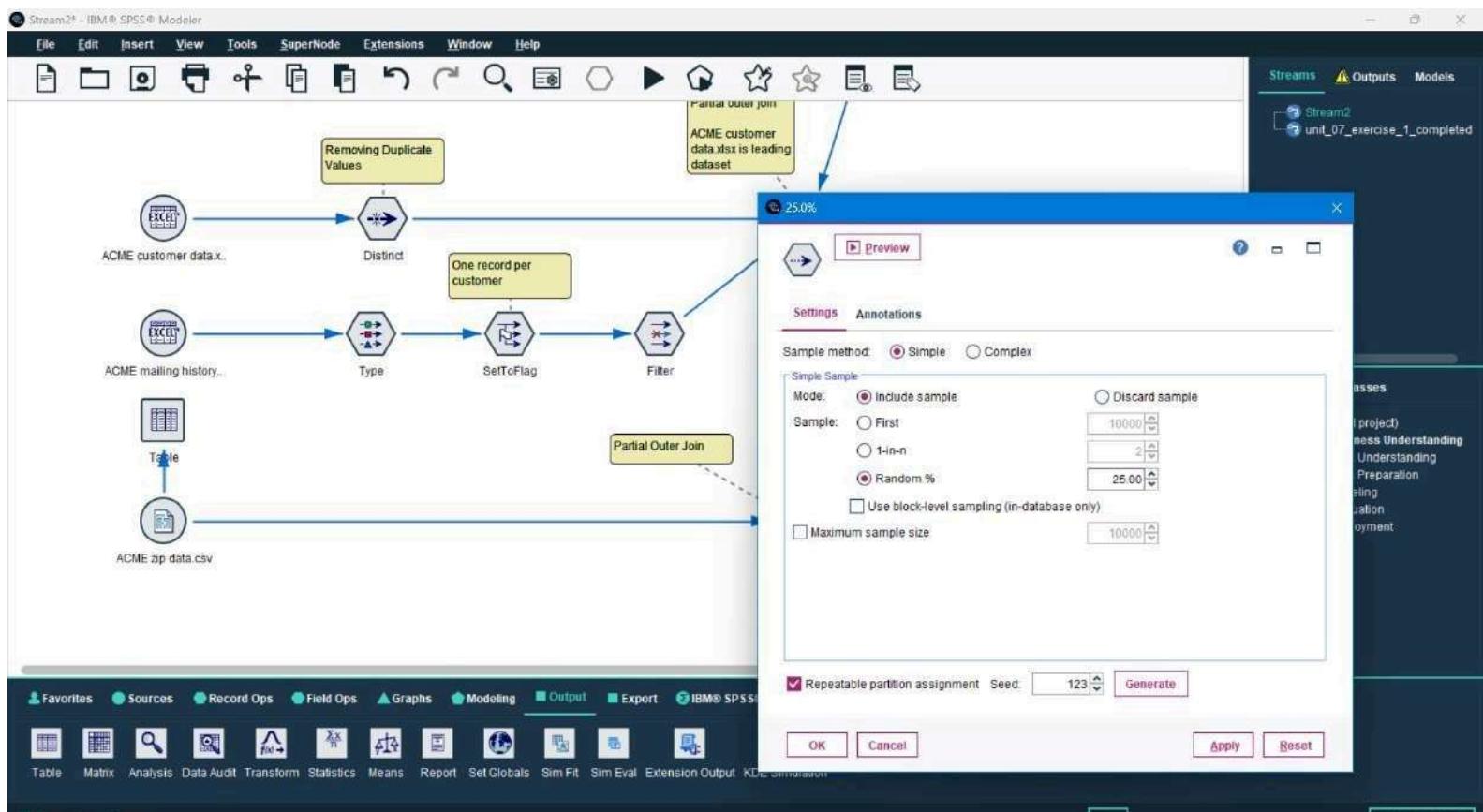
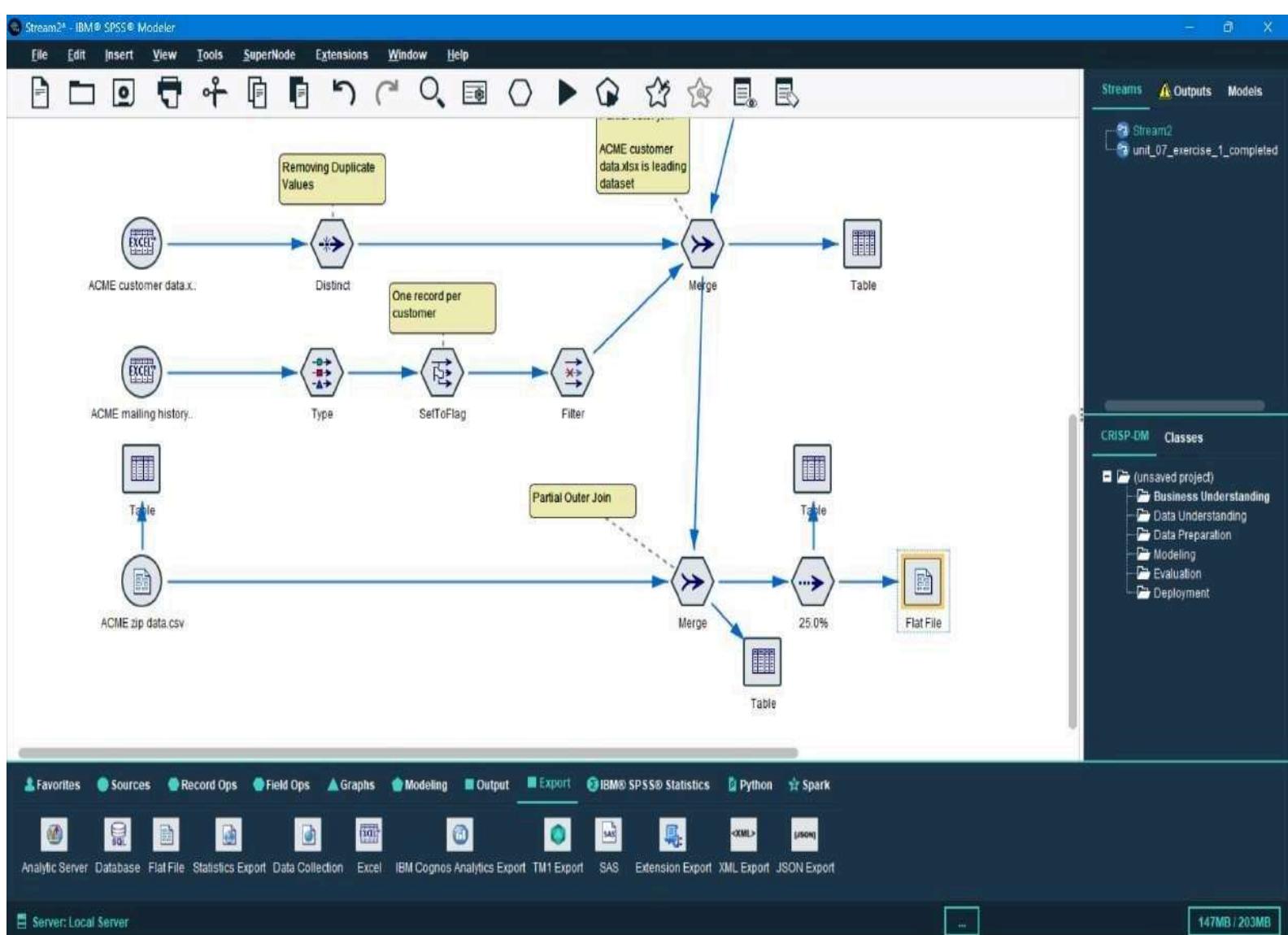


Table (15 fields, 7,560 records)

File Edit Generate

Annotations

	ZIP	pct_working	pct_house_owners	pct_with_car	CUSTOMER_ID	ORDERDATE_Max	PRICE_Sum_Sum	NUMBER OF PURCHASES	GENDER	CREDITLIMIT	ZODIAC	E-MAIL_ADDRESS	mailing_Standard	tennisracket	mailing_XL	original or
1	1001FA	76	60	88	1952.000	2009-05-02	244.742	2	F	6747.000	9.000	name25794@wwmail.uk	F	T		
2	1001FA	76	60	88	26952.000	2008-04-20	2811.882	6	F	6747.000	5.000	name1012@wwmail.inc	\$null\$	\$null\$		
3	1001GB	70	48	88	8110.000	2010-12-06	1472.307	6	M	706.000	10.000	name12835@net.uk	\$null\$	\$null\$		
4	1001GB	70	48	88	18110.000	2009-05-23	755.751	2	M	706.000	6.000	name27443@zigzag.be	\$null\$	\$null\$		
5	1001GB	70	48	88	2811.000	2004-08-02	2138.859	2	M	706.000	6.000	name13572@net.uk	F	T		
6	1001PJ	76	59	90	30259.000	2010-12-07	1289.636	3	F	2831.000	1.000	name14640@wwmail.de	\$null\$	\$null\$		
7	1001PR	86	69	91	3092.000	2000-04-27	414.060	1	M	8171.000	6.000	name26954@wwmail.uk	F	T		
8	1001PR	86	69	91	13293.000	2010-07-15	1656.827	6	M	8171.000	9.000	name8830@net.inc	F	T		
9	1002AI	88	72	92	5597.000	1999-06-27	1176.480	2	M	2198.000	3.000	name13167@net.uk	F	T		
10	1002EH	76	59	92	16891.000	2010-10-25	2758.840	5	M	8927.000	3.000	name17381@wwmail.de	T	F		
11	1002EH	76	59	92	16980.000	2002-01-21	847.100	2	M	8927.000	11.000	name28295@zigzag.be	F	T		
12	1002EH	76	59	92	21891.000	2009-01-02	560.442	2	M	8927.000	11.000	name8832@net.inc	\$null\$	\$null\$		
13	1002GZ	79	67	91	18911.000	2009-12-11	1491.175	3	F	8464.000	10.000	name19028@wwmail.es	T	F		
14	1002JX	80	64	92	13158.000	2010-07-11	356.046	2	F	2027.000	5.000	name12186@net.uk	F	T		
15	1002JX	80	64	92	28151.000	2006-02-25	130.775	1	F	2027.000	3.000	name5049@net.jp	F	T		
16	1002JX	80	64	92	23158.000	2011-06-08	8497.062	3	F	2027.000	3.000	name17913@wwmail.es	F	T		
17	1002MP	83	67	91	2449.000	2003-12-01	812.440	1	F	7455.000	8.000	name13628@net.uk	T	T		
18	1002VW	83	70	92	5957.000	2008-12-12	339.055	2	M	4023.000	1.000	name24484@wwmail.org	\$null\$	\$null\$		
19	1002WX	80	67	92	9512.000	2008-09-17	548.202	1	F	5913.000	5.000	name2970@moble.cat	\$null\$	\$null\$		
20	1002WX	80	67	92	4514.000	2006-03-01	1100.822	1	F	5913.000	1.000	name2724@hmail.uk	F	T		
21	1002YB	73	57	93	3101.000	1999-09-04	38.100	1	M	136.000	10.000	name19903@wwmail.es	F	T		
22	1002YB	73	57	93	13302.000	2010-03-27	942.129	2	M	136.000	2.000	name2266@lomejores	\$null\$	\$null\$		
23	1003XJ	71	56	89	7642.000	2005-07-04	877.972	1	M	9136.000	3.000	name17088@wwmail.de	F			
24	1003XJ	71	56	89	12642.000	2006-05-08	1031.727	2	M	9136.000	3.000	name12254@net.uk	\$null\$	\$null\$		
25	1003XJ	71	56	89	27642.000	2010-03-04	1116.407	4	M	9136.000	7.000	name17852@wwmail.es	\$null\$	\$null\$		
26	1003TK	91	72	92	18266.000	2005-09-15	222.172	1	M	5269.000	7.000	name21817@wwmail.inc	I	F		
27	1003TK	91	72	92	8266.000	2010-02-28	287.227	2	M	5269.000	9.000	name23117@wwmail.inc	F	F		
28	1003TK	91	72	92	23266.000	2008-04-29	823.263	3	M	5269.000	12.000	name10529@netinc	T	F		
29	1003UB	67	49	93	15764.000	2010-02-19	1243.295	3	F	8751.000	7.000	name4198@net.jp	\$null\$	\$null\$		
30	1003UB	67	49	93	25764.000	2010-03-02	834.319	3	F	8751.000	5.000	name20269@wwmail.es	T	F		
31	1004CR	80	60	88	15568.000	2010-01-22	1319.576	4	M	9259.000	2.000	name3427@moble.cat	T	T		
32	1004CR	80	60	88	10566.000	2007-09-27	130.227	1	M	9259.000	5.000	name12510@net.uk	F	T		
33	1004CR	80	60	88	560.000	2008-04-03	340.165	2	M	9259.000	4.000	name13164@net.uk	T	T		
34	1004CR	80	60	88	2556.000	2008-10-15	57.389	1	M	9259.000	3.000	name24613@wwmail.org	F	T		
35	1004B	81	65	89	2383.000	2008-09-14	554.697	3	M	7538.000	6.000	name8802@netinc	\$null\$	\$null\$		
36	1004PO	84	70	94	3077.000	2001-05-18	2138.720	2	F	1366.000	11.000	name8467@net.fr	F	T		
37	1004PO	84	70	94	23278.000	2008-08-02	826.559	4	F	1366.000	1.000	name15764@wwmail.de	F	T		
38	1004XV	85	70	88	4770.000	2000-09-17	870.900	2	F	5368.000	9.000	name14847@wwmail.de	F	T		
39	1004XV	85	70	88	24266.000	2010-11-11	975.799	2	F	5368.000	8.000	name676@hmail.us	\$null\$	\$null\$		
40	1004XV	85	70	88	29266.000	2008-09-16	101.945	1	F	5368.000	10.000	name25802@wwmail.uk	\$null\$	\$null\$		
41	1005GB	81	64	87	7542.000	2008-08-05	205.077	2	M	2214.000	2.000	name17288@wwmail.de	\$null\$	\$null\$		
42	1005GR	77	62	88	950.000	2003-12-15	360.240	2	F	8127.000	6.000	name5396@net.jp	F	T		
43	1005GR	70	56	96	4043.000	2007-10-15	R27.694	2	F	4355.000	8.000	nameR202@net.fr	F	T		



PROJECT-6

Derving and Reclassifying the Fields

unit 8 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream1
unit 8

Fields - 15 Records - 30000

Table

ACME data part 1.dat

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Modeling Output Export IBM® SPSS® Statistics Python Spark

Analytic Server Database Var. File Fixed File Statistics File Data Collection IBM Cognos Analytics TM1 Import TWC Import SAS File Excel XML User Input Sim Gen Extension Import Data View JSON Geospatial

Server: Local Server 131MB / 182MB

unit 8 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode

Table (15 fields, 30,000 records) #3

File Edit Generate

Table Annotations

ZIP	CUSTOMER_ID	GENDER	CREDITLIMIT	ZODIAC	E-MAIL ADDRESS	ORDERD...	AMOUNT_SPENT
1 1818BD	723.000 F		9026.000	1.000	name7502@wmail.fr	2009-07-02	546.731
2 1132DG	724.000 M		5223.000	3.000	name25495@wmail.org	1999-07-11	306.480
3 1803YT	725.000 F		668.000	8.000	name15543@wmail.de	2007-02-27	1796.016
4 1205WR	726.000 M		8302.000	6.000	name28335@igcag.be	2010-06-14	2489.259
5 1711ON	727.000 F		997.000	8.000	name5354@inet.jp	1999-02-21	377.210
6 1055FG	728.000 M		1955.000	3.000	name20637@wmail.es	2004-05-22	624.015
7 1254MR	729.000 F		6898.000	5.000	name20636@wmail.es	2006-07-13	34.724
8 1723DG	730.000 F		6488.000	11.000	name10414@inet.ln	2002-11-29	788.150
9 1713AQ	731.000 F		5006.000	2.000	name23372@wmail.inc	2003-07-17	2419.550
10 1264EC	732.000 F		929.000	5.000	name20635@wmail.es	2003-05-10	1158.180
11 1648BT	733.000 F		3303.000	8.000	name5356@inet.jp	2003-12-01	1247.970
12 1285XV	734.000 F		8934.000	3.000	name17562@wmail.de	2010-07-28	1036.026
13 1282NB	735.000 F		9808.000	7.000	name6386@inet.fr	2009-05-26	457.036
14 1799IT	736.000 M		4723.000	6.000	name10409@inet.ln	2003-10-22	2119.110
15 1802UO	737.000 M		672.000	7.000	name13849@inet.uk	2003-03-09	544.040
16 1971NK	738.000 F		7801.000	8.000	name25473@wmail.org	2002-11-13	406.510
17 1361RL	739.000 M		5984.000	5.000	name13848@inet.uk	2003-01-28	911.980
18 1164VN	740.000 F		2621.000	9.000	name23366@wmail.inc	2001-02-05	1971.720
19 1757YN	741.000 M		6062.000	6.000	name3188@molbe.cat	2001-10-04	1316.750
20 1681HP	742.000 M		9797.000	4.000	name1606@lomejores	2000-01-22	386.320

OK

Streams Outputs Models

Stream1
unit 8

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Modeling Output Export IBM® SPSS® Statistics Python Spark

Analytic Server Database Var. File Fixed File Statistics File Data Collection IBM Cognos Analytics TM1 Import TWC Import SAS File Excel XML User Input Sim Gen Extension Import Data View JSON Geospatial

Server: Local Server 132MB / 182MB

unit 8 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream1 unit 8

Fields - 16
Records - 30000
Field Added -
DIFFERENCE_CL_AS

Fields - 15
Records - 30000

Table Table

ACME data part 1.dat DIFFERENCE_CL_AS

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Var. File Fixed File Statistics File Data Collection IBM Cognos Analytics TM1 Import TWC Import SAS File Excel XML User Input Sim Gen Extension Import Data View JSON Geospatial

Server: Local Server 134MB / 183MB

```
graph TD; ACME[ACME data part 1.dat] --> Table1[Table]; ACME --> Table2[Table]; Table1 --> DIFFERENCE[DIFERENCE_CL_AS]; Table2 --> DIFFERENCE;
```

unit 8 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

DIFFERENCE_CL_AS

Preview

Settings Annotations

Mode: Single Multiple

Derive field:
DIFFERENCE_CL_AS

Derive as: Formula

Field type: <Default>

Formula:
1 CREDITLIMIT - AMOUNT_SPENT

OK Cancel Apply Reset

Streams Outputs Models

Stream1 unit 8

CRISP-DM Classes

(unsaved project)

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM SPSS Statistics Python Spark

Analytic Server Database Var. File Fixed File Statistics File Data Collection IBM Cognos Analytics TM1 Import TWC Import SAS File Excel XML User Input Sim Gen Extension Import Data View JSON Geospatial

Server: Local Server 135MB / 183MB

```
graph TD; ACME[ACME data part 1.dat] --> Table1[Table]; ACME --> Table2[Table]; Table1 --> DIFFERENCE[DIFERENCE_CL_AS]; Table2 --> DIFFERENCE;
```

Table (16 fields, 30,000 records) #4

File Edit Generate

Table Annotations

P	CUSTOMER_ID	GENDER	CREDITLIMIT	ZODIAC	E-MAIL ADDRESS	ORDERID	AMOUNT_SPENT	MAILING_ID_1.0	MAILING_ID_2.0	1.0_RESPONSE	2.0_RESPONSE	pct_working	pct_with_car	pct_house_owners	DIFFERENCE	CL_AS
1	116BO	723.000 F	9026.000	1.000	name7502@net.fr	2009-07-02	548.731 F	T		0.000	0.000	82	88	66	8479.269	
2	132DG	724.000 M	5223.000	3.000	name25485@wwmail.org	1999-07-11	306.480 F	T		Snnull\$	0.000	82	86	67	4916.520	
3	103YT	725.000 F	668.000	8.000	name1543@wwmail.de	2007-02-27	1795.016 F	T		Snnull\$	0.000	91	92	76	-1128.016	
4	055WR	726.000 M	8302.000	6.000	name2833@zigzag.be	2010-06-14	2489.259 F	T		Snnull\$	0.000	79	85	63	5812.741	
5	711ON	727.000 F	997.000	8.000	name5354@net.jp	1999-02-21	377.210 F	T		Snnull\$	0.000	80	92	64	619.790	
6	755FG	728.000 M	1955.000	3.000	name20637@wwmail.es	2004-05-22	624.015 T	T		0.000	1.000	84	92	71	1330.985	
7	254MR	729.000 F	6898.000	5.000	name20636@wwmail.es	2006-07-13	34.724 F	T		Snnull\$	0.000	73	91	57	6863.276	
8	723DG	730.000 F	6488.000	11.000	name10414@netinc	2002-11-29	788.150 F	T		Snnull\$	0.000	79	92	87	5699.850	
9	713AQ	731.000 F	5006.000	2.000	name23372@wwmail.inc	2003-07-17	2419.550 F	T		Snnull\$	0.000	76	93	61	2586.450	
10	664EC	732.000 F	929.000	5.000	name20035@wwmail.es	2003-05-10	1158.180 F	T		Snnull\$	0.000	80	86	64	-229.180	
11	448BT	733.000 F	3303.000	8.000	name5356@net.jp	2003-12-01	1247.970 T	T		0.000	0.000	80	92	63	2055.030	
12	285XV	734.000 F	8934.000	3.000	name17582@wwmail.de	2010-07-28	1038.025 F	T		Snnull\$	0.000	83	91	68	7897.974	
13	982NB	735.000 F	9808.000	7.000	name6388@net.fr	2009-05-26	457.035 F	T		Snnull\$	0.000	77	92	63	9350.964	
14	998IT	736.000 M	4723.000	6.000	name10409@netinc	2003-10-22	2119.110 F	T		Snnull\$	0.000	72	92	56	2603.890	
15	102UO	737.000 M	672.000	7.000	name13849@netuk	2003-03-09	544.040 F	T		Snnull\$	0.000	85	89	87	127.960	
16	771NK	738.000 F	7801.000	8.000	name25473@wwmail.org	2002-11-13	406.510 T	T		0.000	0.000	81	90	70	7394.490	
17	611RL	739.000 M	5984.000	5.000	name13848@net.uk	2003-01-28	911.980 F	T		Snnull\$	0.000	77	92	62	5072.020	
18	164VN	740.000 F	2621.000	9.000	name23366@wwmail.in	2001-02-05	1971.720 F	T		Snnull\$	0.000	80	89	64	649.280	
19	677YN	741.000 M	6062.000	6.000	name3188@mobcat	2010-10-04	1316.750 F	T		Snnull\$	0.000	83	86	67	4745.250	
20	581HP	742.000 M	9797.000	4.000	name1605@omejor.es	2000-01-22	386.320 F	T		Snnull\$	0.000	70	88	53	9410.680	
21	787PI	743.000 F	2687.000	6.000	name29056@zigzag.cat	2003-09-16	1454.310 T	T		0.000	0.000	82	87	64	1232.960	
22	123AR	744.000 M	128.000	9.000	name28323@zigzag.be	2004-08-09	711.319 F	T		Snnull\$	1.000	83	90	88	-583.319	
23	737ZI	745.000 F	182.000	11.000	name20303@wwmail.es	2001-10-20	1199.380 F	T		Snnull\$	0.000	82	88	62	-1017.380	
24	252T2	746.000 M	2301.000	9.000	name13947@net.uk	2008-01-28	165.638 F	T		Snnull\$	0.000	82	91	63	2135.382	
25	684YE	747.000 F	2704.000	8.000	name10401@netinc	2008-08-15	207.407 F	T		Snnull\$	0.000	85	87	69	2496.593	
26	508OU	748.000 M	4033.000	7.000	name13846@net.uk	2003-08-27	785.060 T	T		0.000	0.000	89	90	76	3247.940	
27	143VC	749.000 M	1818.000	9.000	name17579@wwmail.de	2009-02-08	128.067 F	T		Snnull\$	0.000	85	90	71	1689.933	
28	288IK	750.000 F	9044.000	6.000	name26894@wwmail.uk	2009-02-23	1365.367 F	T		Snnull\$	0.000	80	91	60	7076.633	
29	677CK	751.000 F	1975.000	12.000	name5356@net.fr	1999-10-24	2091.560 F	T		Snnull\$	0.000	76	87	59	-116.560	
30	115FI	752.000 F	3619.000	9.000	name1609@omejor.es	2002-12-13	1757.480 F	T		Snnull\$	0.000	86	88	71	1861.520	
31	753OI	753.000 M	3517.000	8.000	name29727@zigzag.cat	2002-01-18	454.800 T	T		0.000	0.000	84	90	67	3062.200	
32	129DC	754.000 M	701.000	8.000	name29850@zigzag.es	2010-04-29	405.287 F	T		Snnull\$	0.000	82	88	67	295.713	
33	036LJ	755.000 M	1878.000	3.000	name17578@wwmail.de	2004-05-14	573.874 F	T		Snnull\$	1.000	78	89	63	1304.126	
34	102VW	756.000 M	4023.000	10.000	name29049@zigzag.cat	2004-11-15	3215.550 F	T		Snnull\$	1.000	83	92	70	807.450	
35	779KZ	757.000 M	4946.000	10.000	name10389@netinc	2000-05-01	959.500 F	T		Snnull\$	0.000	82	87	66	3988.500	
36	784IE	758.000 F	9946.000	8.000	name6354@net.fr	2010-10-30	971.262 T	T		0.000	0.000	82	87	67	8974.738	
37	183FC	759.000 F	1522.000	6.000	name3172@mobcat	2003-03-04	1183.400 F	T		Snnull\$	0.000	85	93	66	338.600	
38	192WX	760.000 F	7458.000	2.000	name433@mall.de	2001-11-19	31.560 F	T		Snnull\$	0.000	74	88	62	7426.440	
39	569VY	761.000 F	1340.000	10.000	name5357@net.jp	2002-11-24	1195.590 F	T		Snnull\$	0.000	87	92	73	144.410	
40	580OG	762.000 F	7136.000	3.000	name25456@wwmail.in	2002-12-10	286.970 F	T		Snnull\$	0.000	94	89	81	6849.030	
41	188SB	763.000 F	1723.000	4.000	name25457@wwmail.org	2011-06-30	1687.274 T	T		1.000	0.000	74	92	64	35.728	
42	172BQ	764.000 F	3799.000	7.000	name1616@omejor.es	1989-09-07	845.380 F	T		Snnull\$	0.000	73	90	60	2953.620	
43	1nour	765.000 M	1226.000	2.000	name2402@netin.in	2007-02-10	814.531 F	T		Snnull\$	0.000	97	92	60	507.479	

unit 8* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

```

graph LR
    A[Fields - 15  
Records - 30000] --> B[Fields - 16  
Records - 30000  
Field Added - DIFFERENCE_CL_AS]
    C[Fields - 19  
Records - 30000  
Fields Added - CREDITLIMIT_DOLLAR  
AMOUNT_SPENT_DOLLAR  
DIFFERENCE_CL_AS_DOLLAR] --> D[Difference_Cl_As]
    E[Difference_Cl_As] --> F[Dollar]
    
```

Streams Outputs Models

CRISP-DM Classes

- (unavailable project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Favorites Sources Record Ops Field Ops Modeling Output Export IBM® SPSS® Statistics Python Spark

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation

Server: Local Server 139MB / 183MB

Unit 8 - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream1 unit 8

```

graph LR
    A[Fields - 15  
Records - 30000] --> B[Table]
    C[Fields - 16  
Records - 30000  
Field Added:  
DIFFERENCE_CL_AS] --> D[Table]
    E[Fields - 19  
Records - 30000  
Field Added:  
CREDITLIMIT_DOLLAR  
AMOUNT_SPENT_DOLLAR  
DIFFERENCE_CL_AS_DOLLAR] --> F[Table]
    B --> G{DIFFERENCE_CL_AS}
    D --> G
    F --> G
    G --> H{_Dollar}
  
```

The Stream diagram shows three input tables (ACME data part 1.dat, DIFFERENCE_CL_AS, and _Dollar) being joined into a single output table (_Dollar). The DIFFERENCE_CL_AS node adds a new field, DIFFERENCE_CL_AS, to the joined tables.

_Dollar Derive as: Formula

- Mode: Multiple
- Derive from: CREDITLIMIT, AMOUNT_SPENT, DIFFERENCE_CL_AS
- Field name extension: _Dollar
- Derive as: Formula
- Field type: <Default>
- Formula: @FIELD * 1.4

CRISP-DM Classes: (unsaved project), Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBMSPSS

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globala Sim Fit Sim Eval Extension Out

Server: Local Server 140MB / 183MB

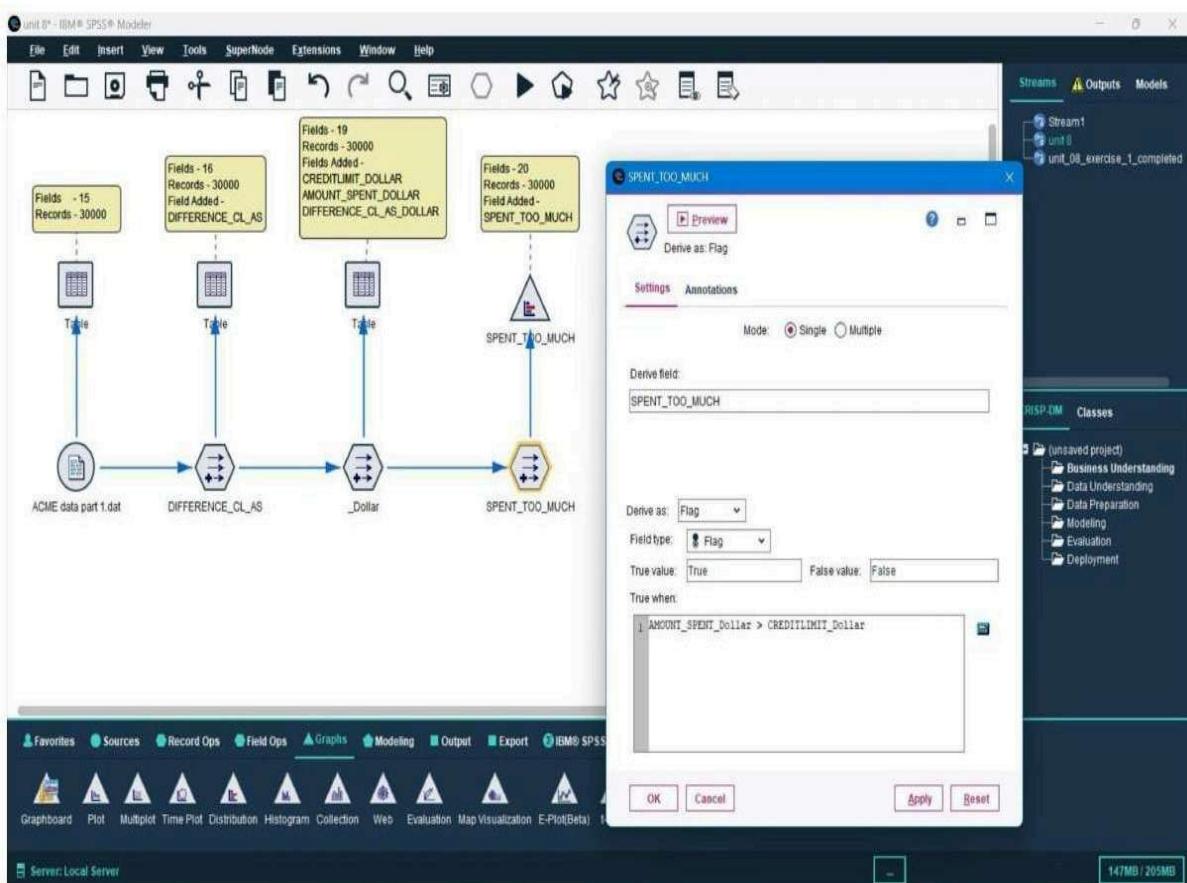
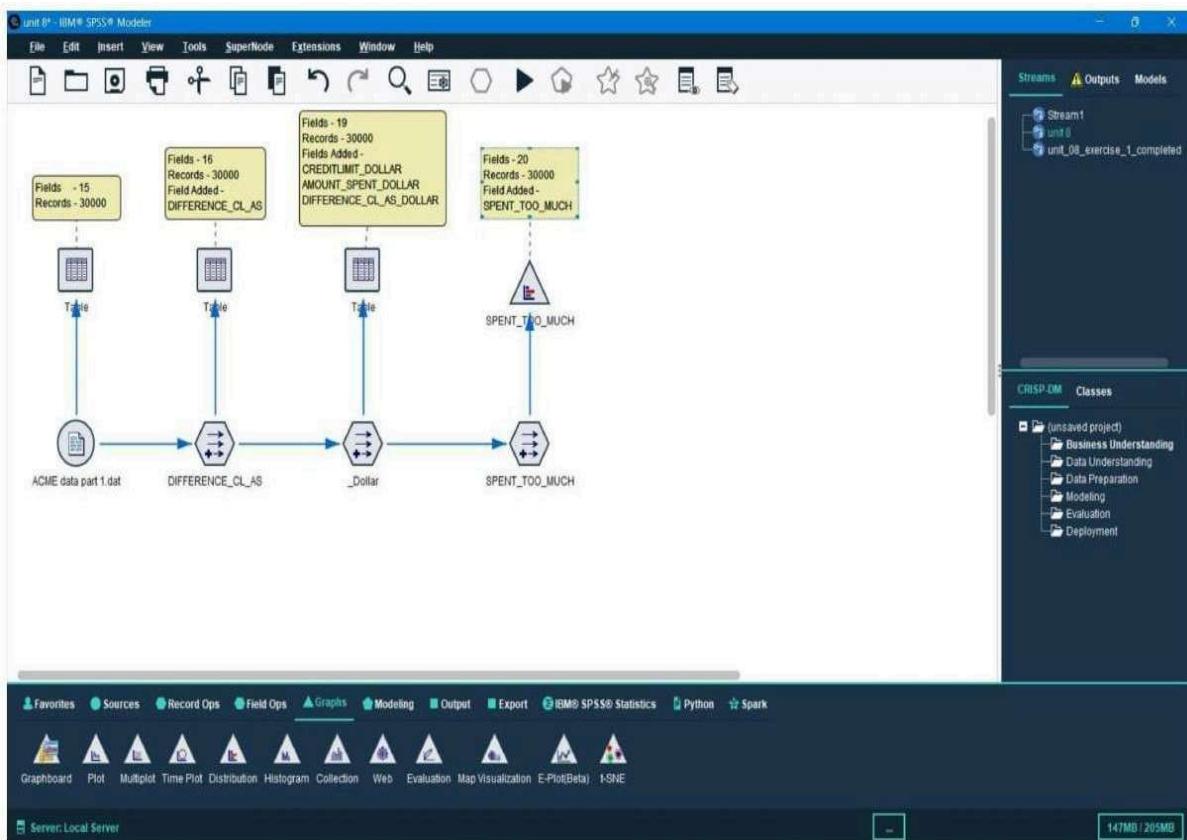
Table (19 fields, 30,000 records) #1

File Edit Generate

Table Annotations

	ORDERID	AMOUNT_SPENT	MAILING_ID_1.0	MAILING_ID_2.0	1.0_RESPONSE	2.0_RESPONSE	pct_working	pct_with_car	pct_house_owners	DIFFERENCE_CL_AS	CREDITLIMIT_Dollar	AMOUNT_SPENT_Dollar	DIFFERENCE_CL_AS_Dollar
1	2009-07-02	548.731T	T		0.000	0.000	82	88	66	8479.269	12836.400	765.424	11870.978
2	mail.org 1999-07-11	306.480F	T		\$null\$	0.000	82	86	67	4916.520	7312.200	429.072	6883.128
3	mail.org 2007-02-27	1796.016F	T		\$null\$	0.000	91	92	76	-1128.016	935.200	2514.422	-1579.222
4	g.be 2010-06-14	2469.259F	T		\$null\$	0.000	79	85	63	5812.741	11622.800	3484.962	8137.838
5	1999-02-21	377.210F	T		\$null\$	0.000	80	92	64	619.790	1395.800	528.094	867.706
6	mail.es 2004-05-22	624.015T	T		0.000	1.000	84	92	71	1330.985	2737.000	873.621	1863.379
7	mail.es 2006-07-13	34.724F	T		\$null\$	0.000	73	91	57	8863.278	9657.200	48.914	9508.588
8	nc 2002-11-29	788.150F	T		\$null\$	0.000	79	92	67	5699.850	9083.200	1103.410	7979.790
9	mail.inc 2003-07-17	2419.550F	T		\$null\$	0.000	76	93	61	2585.450	7008.400	3387.370	3621.030
10	mail.es 2003-05-10	1158.180F	T		\$null\$	0.000	80	86	64	-229.180	1300.600	1621.452	-320.852
11	2003-12-01	1247.970T	T		0.000	0.000	80	92	63	2055.030	4824.200	1747.158	2877.042
12	mail.de 2010-07-28	1036.026F	T		\$null\$	0.000	83	91	68	7897.974	12507.600	1450.437	11057.163
13	2009-05-26	457.036F	T		\$null\$	0.000	77	92	63	9350.954	13731.200	639.851	13091.349
14	nc 2003-10-22	2119.110F	T		\$null\$	0.000	72	92	55	2603.890	6612.200	2986.754	3645.446
15	uk 2003-03-09	544.040F	T		\$null\$	0.000	85	89	67	127.950	940.800	761.656	179.144
16	mail.org 2002-11-13	406.510T	T		0.000	0.000	81	90	70	7394.490	10921.400	589.114	10352.288
17	uk 2003-01-28	911.980F	T		\$null\$	0.000	77	92	62	5072.020	8377.600	1276.772	7100.928
18	mail.inc 2001-02-05	1971.720F	T		\$null\$	0.000	80	89	64	649.280	3669.400	2760.408	908.992
19	ca 2001-10-04	1316.750F	T		\$null\$	0.000	83	86	67	4745.250	8486.800	1843.450	6643.350
20	ores 2000-01-22	386.320F	T		\$null\$	0.000	70	88	53	9410.680	13715.800	540.848	13174.952
21	g.cat 2003-09-16	1454.310T	T		0.000	0.000	82	87	64	1232.690	3761.800	2036.034	1725.766
22	g.be 2004-08-09	711.319F	T		\$null\$	1.000	83	90	68	-583.319	179.200	995.847	-816.647
23	mail.es 2001-10-20	1199.380F	T		\$null\$	0.000	82	88	62	-1017.380	254.800	1679.132	-1424.332
24	uk 2008-01-28	165.638F	T		\$null\$	0.000	82	91	63	2135.362	3221.400	231.893	2989.507
25	nc 2008-08-15	207.407F	T		\$null\$	0.000	85	87	69	2496.593	3785.600	290.369	3495.231
26	uk 2003-08-27	785.060T	T		0.000	0.000	89	90	76	3247.940	5646.200	1099.084	4547.116
27	mail.de 2009-02-08	128.067F	T		\$null\$	0.000	85	90	71	1689.933	2545.200	179.294	2365.906
28	mail.uk 2009-02-23	1365.367F	T		\$null\$	0.000	80	91	60	7587.833	12681.600	1911.514	10750.098
29	1999-10-24	2091.560F	T		\$null\$	0.000	76	87	59	-116.560	2765.000	2928.184	-163.184
30	ores 2002-12-13	1757.480F	T		\$null\$	0.000	85	88	71	1861.520	5066.600	2460.472	2906.128
31	g.com 2002-01-18	454.800T	T		0.000	0.000	84	90	57	3062.200	4923.800	635.720	4287.080
32	ges 2010-04-29	405.287F	T		\$null\$	0.000	82	88	67	295.713	981.400	567.401	413.999
33	mail.de 2004-05-14	573.874F	T		\$null\$	1.000	78	89	63	1304.126	2629.200	803.424	1825.776
34	g.cat 2004-11-15	3215.550F	T		\$null\$	1.000	83	92	70	807.450	5632.200	4501.769	1130.431
35	nc 2000-05-01	959.500F	T		\$null\$	0.000	82	87	66	3986.500	6924.400	1343.300	5581.100
36	2010-10-30	971.282T	T		0.000	0.000	82	87	67	8974.738	13924.400	1389.767	12564.633
37	cat 2003-03-04	1183.400F	T		\$null\$	0.000	85	93	66	338.500	2130.800	1656.760	474.040
38	2001-11-19	31.560F	T		\$null\$	0.000	74	88	62	7426.440	10441.200	44.184	10397.016
39	2002-11-24	1195.590F	T		\$null\$	0.000	87	92	73	144.410	1876.000	1673.826	202.174
40	ic 2002-12-10	286.970F	T		\$null\$	0.000	94	89	81	6849.030	9990.400	401.758	9588.642
41	mail.org 2011-06-30	1687.274T	T		1.000	0.000	74	92	64	35.726	2412.200	2362.183	50.017
42	ores 1999-09-07	845.380F	T		\$null\$	0.000	73	90	60	2953.620	5318.600	1183.532	4135.068
43	2007-07-10	818.530F	T		\$null\$	n nnn	87	97	69	4017.477	1894.400	1146.942	710.458

OK



Unit 08 - IBM® SPSS® Modeler

The screenshot shows the IBM SPSS Modeler interface. On the left, a stream diagram is displayed with three parallel paths. The first path starts with a 'Table' node containing 15 fields and 30,000 records, which is connected to a 'Join' node labeled 'DIFERENCE_CL_AS'. The second path starts with a 'Table' node containing 16 fields and 30,000 records, with a new field 'FIELD_ADDED - DIFFERENCE_CL_AS' added, which is also connected to the 'Join' node. The third path starts with a 'Table' node containing 19 fields and 30,000 records, with new fields 'CREDITLIMIT_DOLLAR' and 'AMOUNT_SPENT_DOLLAR' added, followed by a 'Join' node labeled '_Dollar'. The output of these three paths is connected to a 'SPEN...' node. A 'SPENT_TOO MUCH' dialog box is open in the center, showing a histogram of the 'SPENT_TOO MUCH' field. The dialog has tabs for Plot, Appearance, Output, and Annotations, with the Plot tab selected. Under Plot, 'Selected fields' is chosen, and the field 'SPENT_TOO MUCH' is selected. The 'Color' dropdown is empty, and the 'Normalize by color' checkbox is unchecked. The 'Sort' section has 'Alphabetic' selected. A checked checkbox at the bottom indicates 'Proportional scale'. At the bottom of the dialog are buttons for OK, Run, Cancel, Apply, and Reset.

Streams Outputs Models

Stream1
unit 08
unit_08_exercise_1_completed

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

ACME data part 1.dat DIFFERENCE_CL_AS _Dollar SPEN...

Plot Appearance Output Annotations

Plot: Selectedfields All flags (true values)

Field: SPENT_TOO MUCH

Overlay:

Color:

Normalize by color

Sort: Alphabetic By count

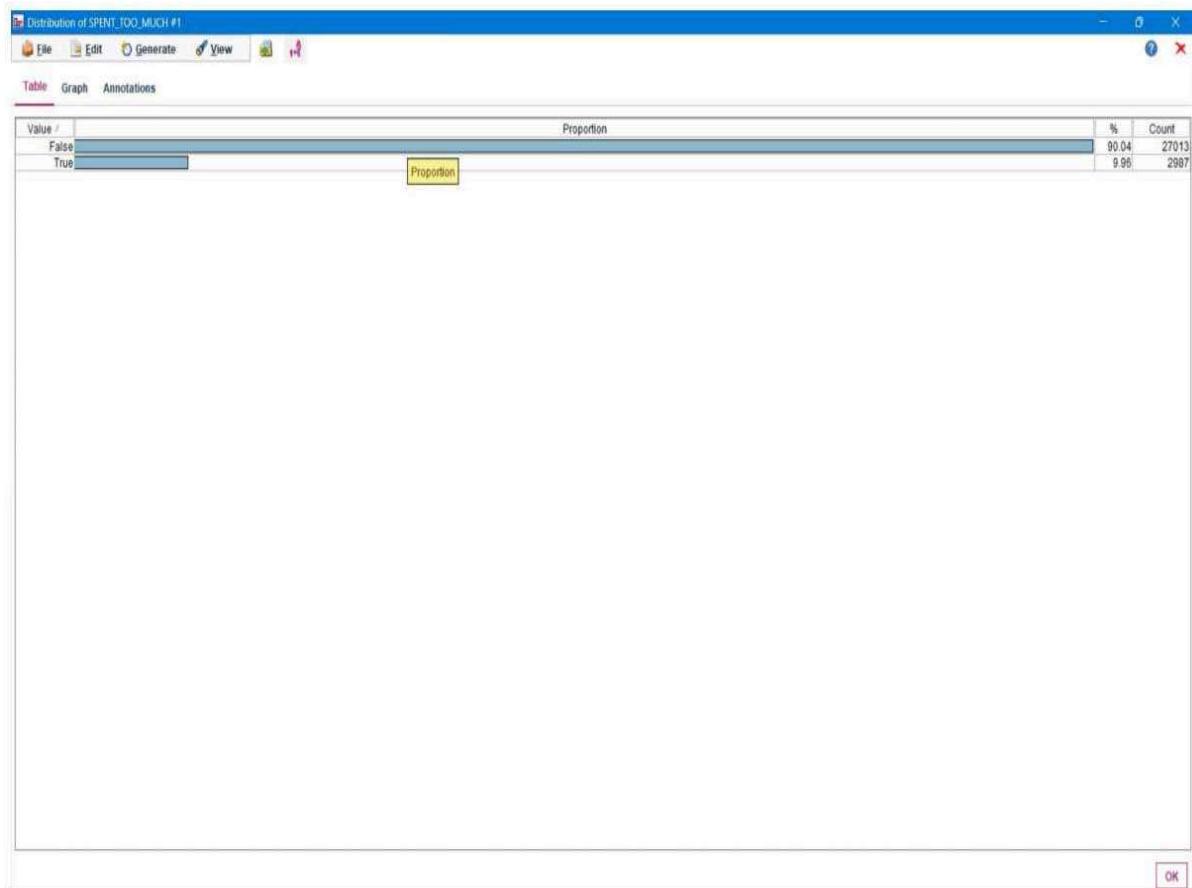
Proportional scale

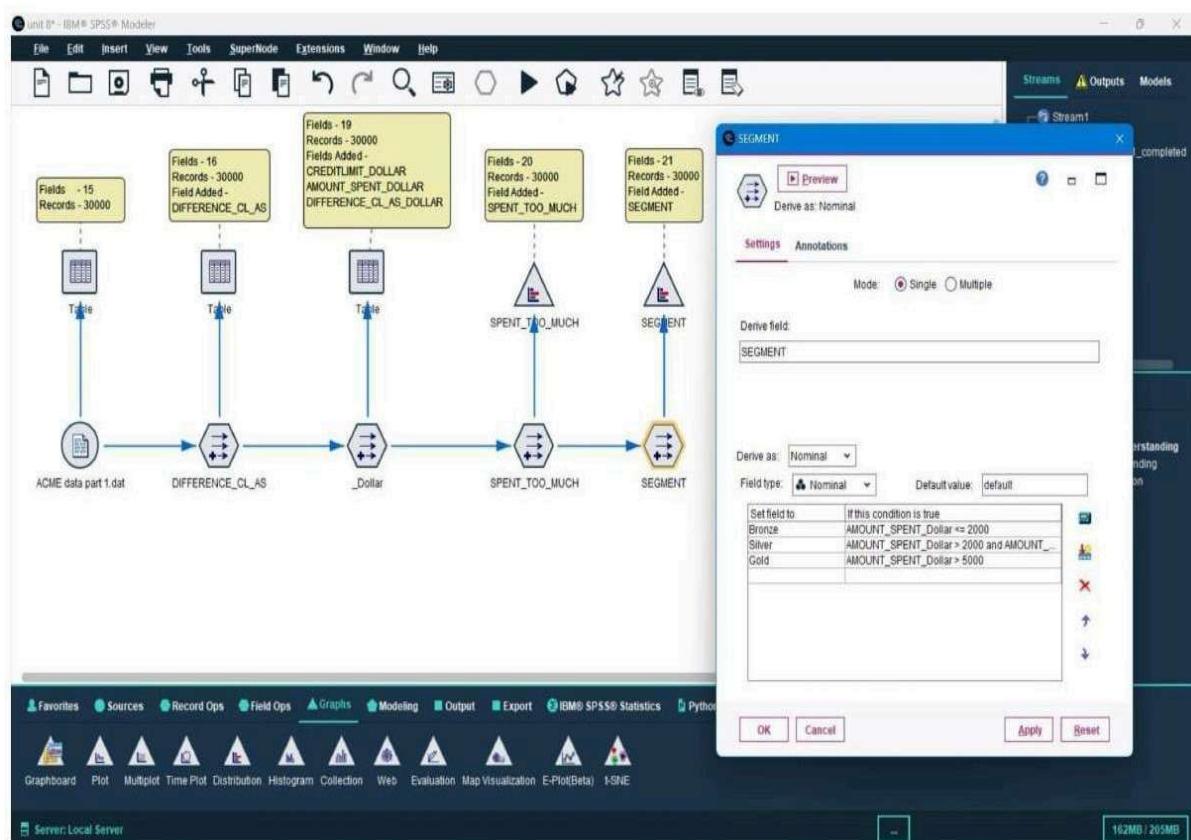
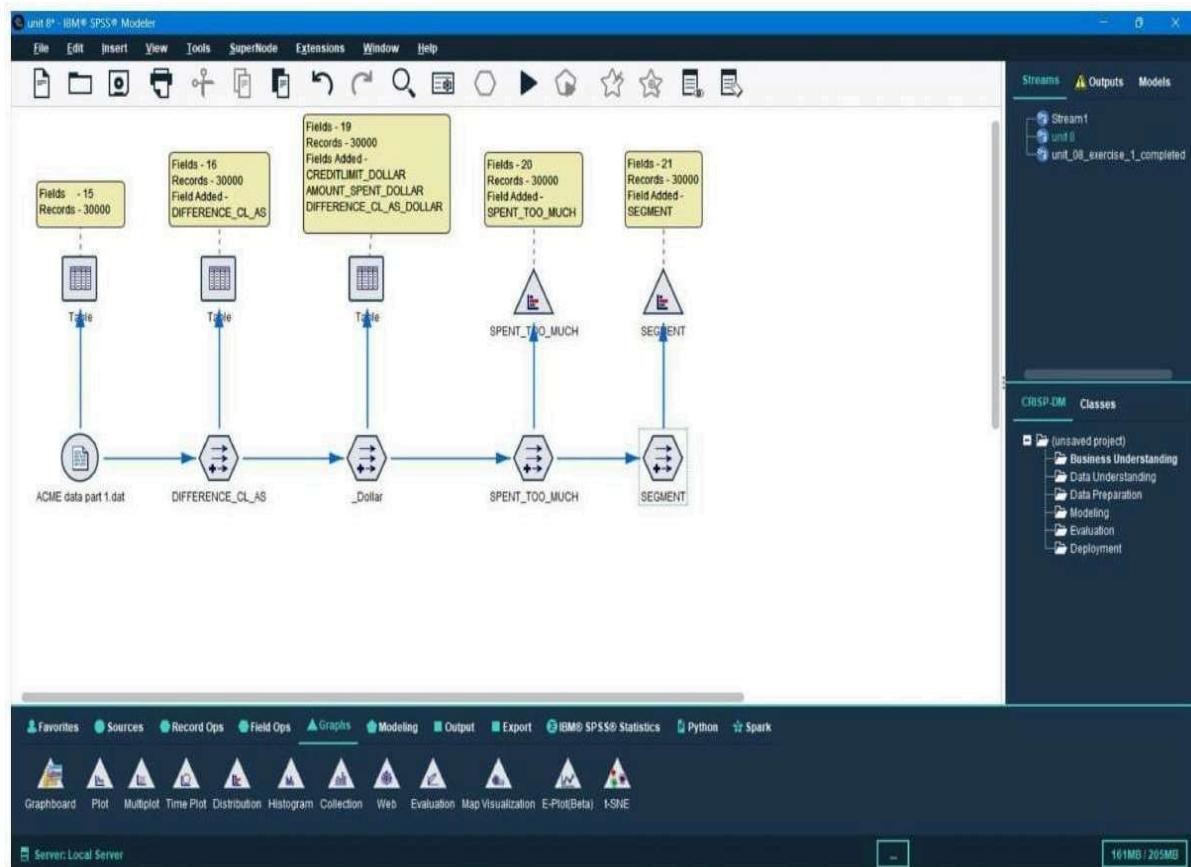
OK Run Cancel Apply Reset

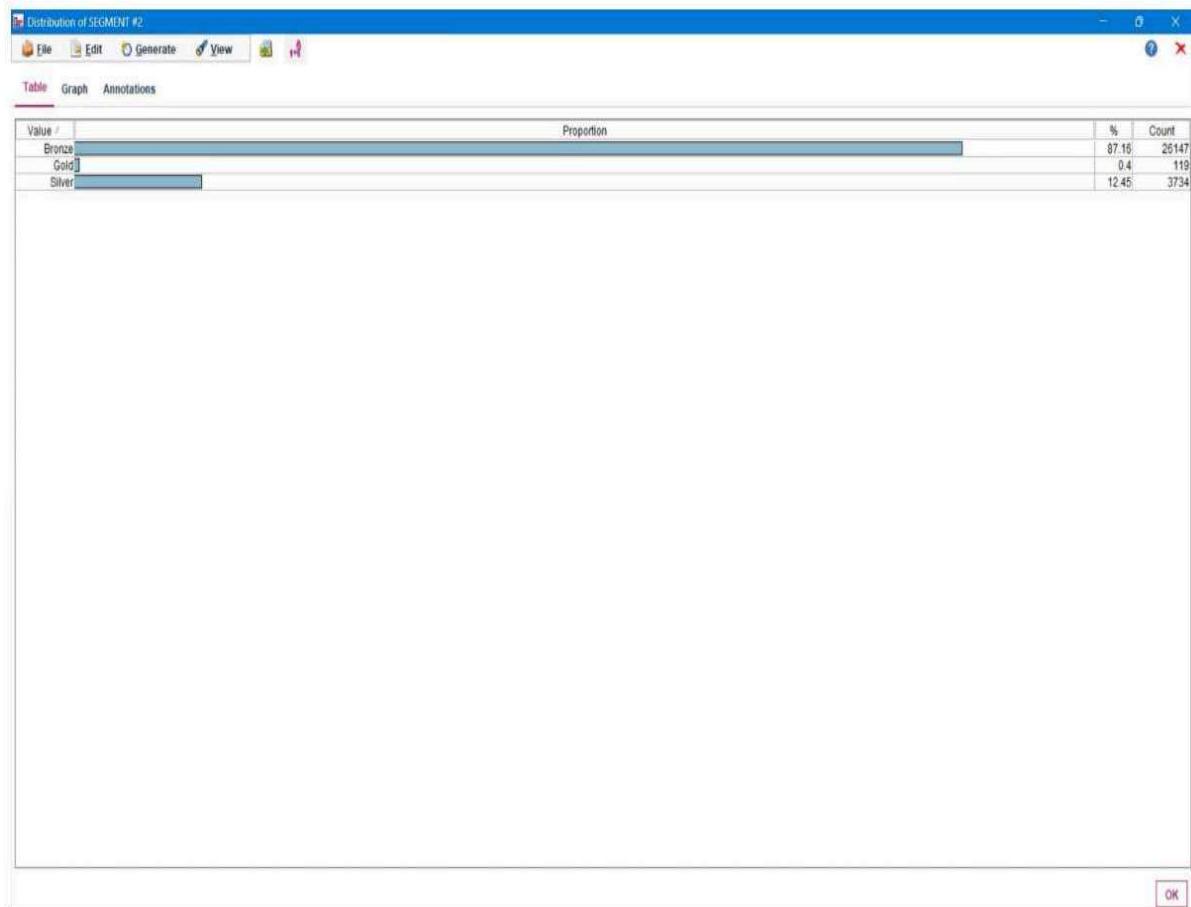
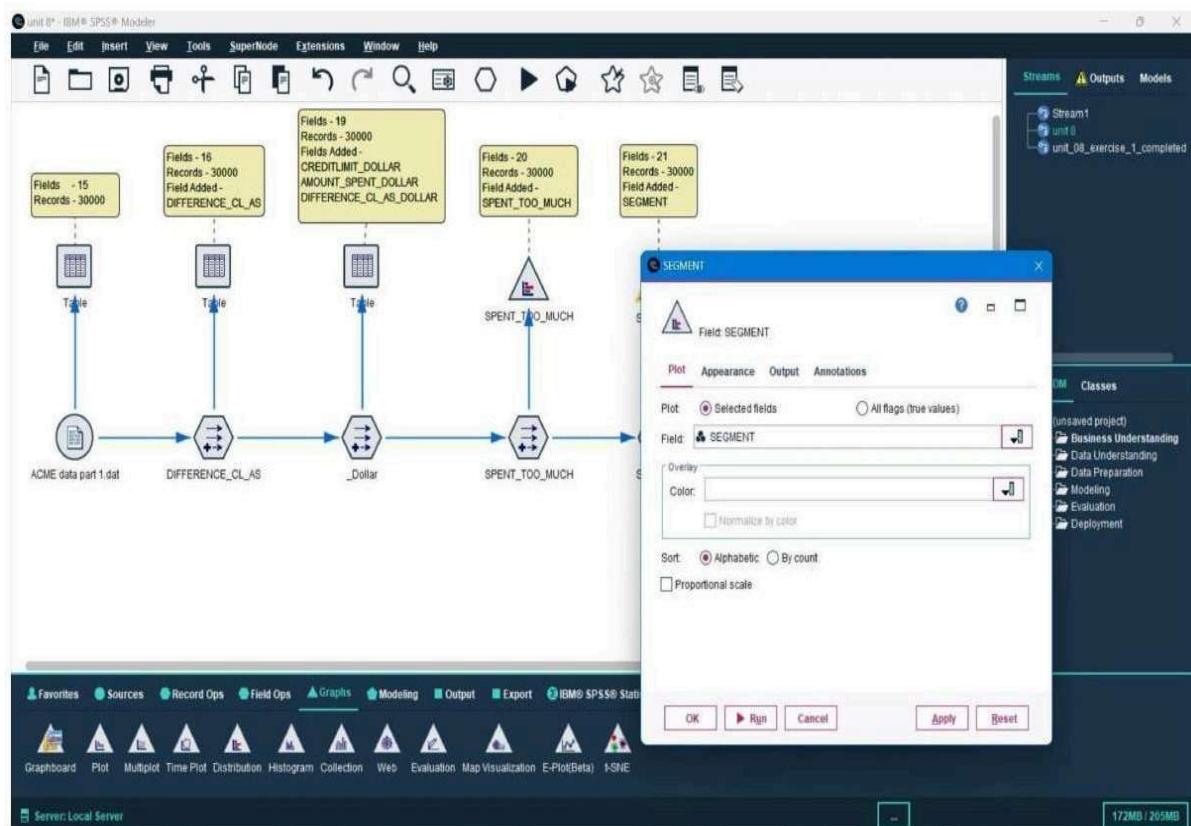
Favorites Sources Record Ops Field Ops Graphs Modeling Output Exports

Graphboard Plot Multiplot Time Plot Distribution Histogram Collection Web Evaluation Map Visualization E-Plot(Beta) t-SNE

Server: Local Server 147MB / 205MB







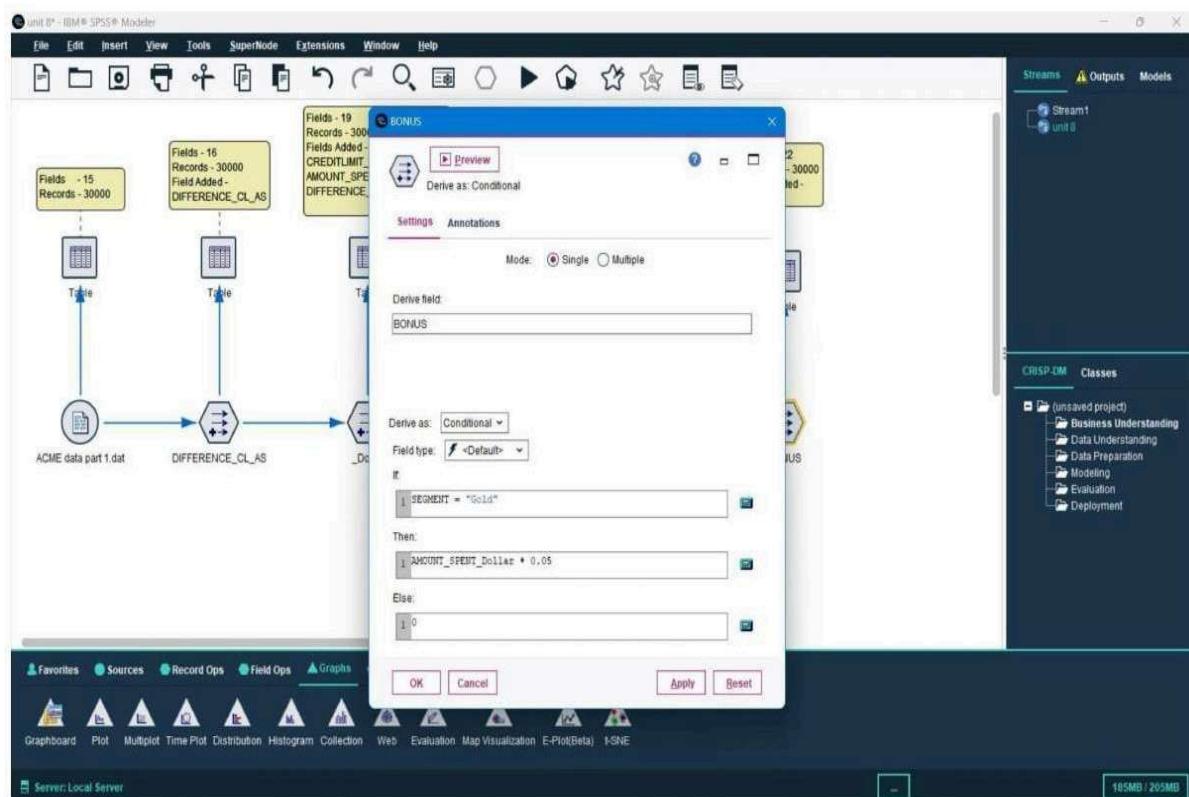
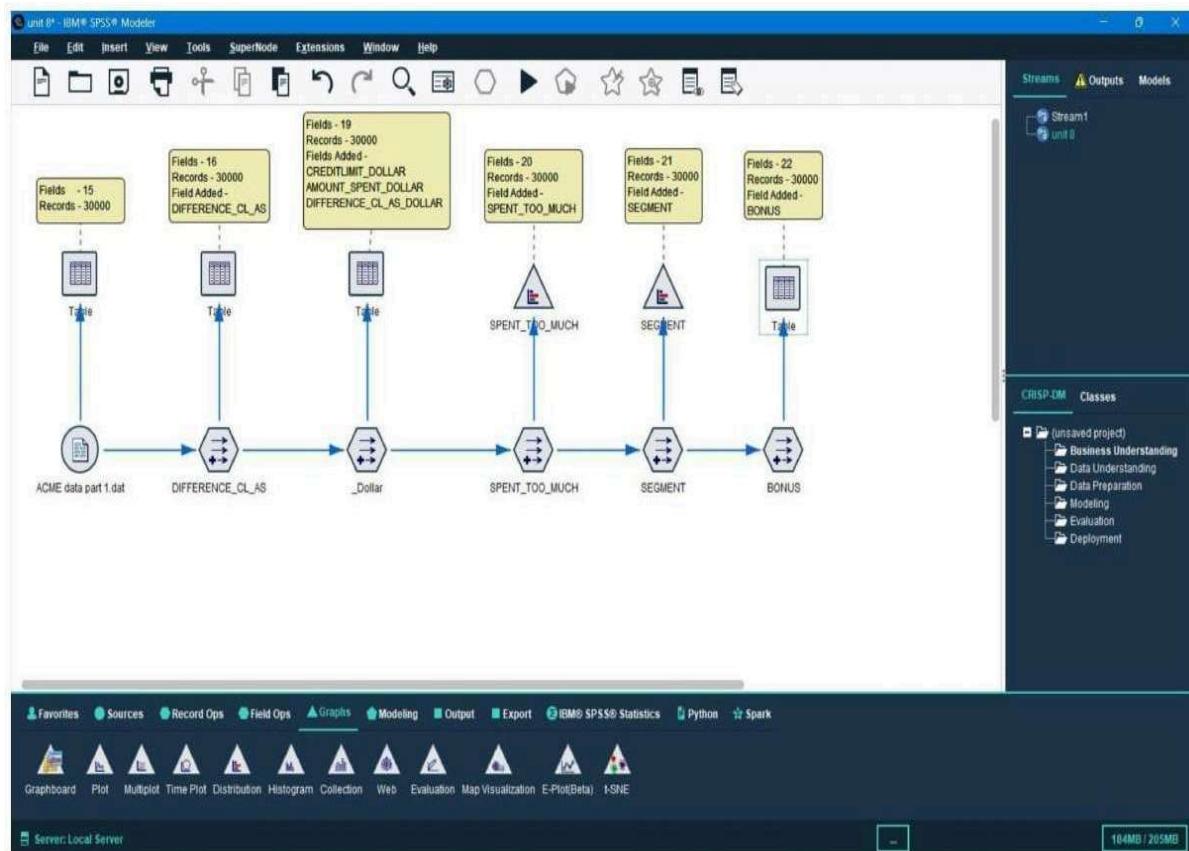


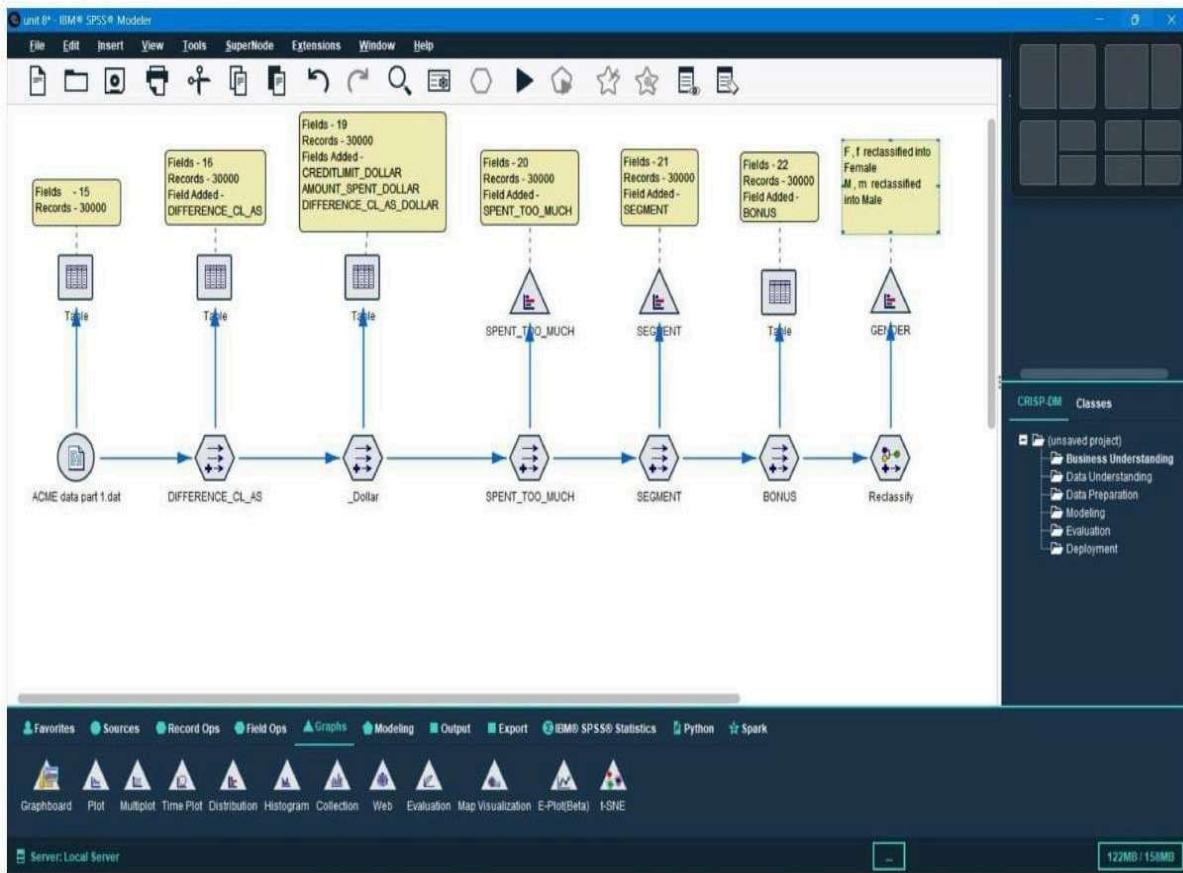
Table (22 fields, 30,000 records) #1

File Edit Generate

Table Annotations

	MAILING_ID_1.0	MAILING_ID_2.0	1.0_RESPONSE	2.0_RESPONSE	pct_working	pct_with_car	pct_house_owners	DIFFERENCE_CL_AS	CREDITLIMIT_Dollar	AMOUNT_SPENT_Dollar	DIFFERENCE_CL_AS_Dollar	SPENT_TOO MUCH	SEGMENT	BONUS
1	T		0.000	0.000	82	88	68	8479.269	12836.400	765.424	11870.978	False	Bronze	0.000
2	T		\$null\$	0.000	82	86	67	4916.520	7312.200	429.072	6883.128	False	Bronze	0.000
3	T		\$null\$	0.000	91	92	76	-1128.016	932.200	2514.422	-1579.222	True	Silver	0.000
4	T		\$null\$	0.000	79	85	63	5812.741	11622.800	3484.962	8137.838	False	Silver	0.000
5	T		\$null\$	0.000	80	92	64	819.790	1395.800	528.094	867.706	False	Bronze	0.000
6	T		0.000	1.000	84	92	71	1330.985	2737.000	873.521	1863.379	False	Bronze	0.000
7	T		\$null\$	0.000	73	91	57	6863.276	9857.200	48.814	9508.586	False	Bronze	0.000
8	T		\$null\$	0.000	79	92	67	5999.950	9083.200	1103.410	7979.790	False	Bronze	0.000
9	T		\$null\$	0.000	76	93	51	2588.450	7008.400	3387.370	3521.030	False	Silver	0.000
10	T		\$null\$	0.000	80	88	64	-229.180	1300.600	1621.452	-320.852	True	Bronze	0.000
11	T		0.000	0.000	80	92	63	2055.030	4624.200	1747.158	2877.042	False	Bronze	0.000
12	T		\$null\$	0.000	83	91	68	7897.974	12507.600	1450.437	11057.163	False	Bronze	0.000
13	T		\$null\$	0.000	77	92	63	9350.964	13731.200	838.851	13091.349	False	Bronze	0.000
14	T		\$null\$	0.000	72	92	55	2603.890	6812.200	2968.754	3645.446	False	Silver	0.000
15	T		\$null\$	0.000	85	89	67	127.960	840.800	7618.586	179.144	False	Bronze	0.000
16	T		0.000	0.000	81	90	70	7394.490	10921.400	589.114	10352.286	False	Bronze	0.000
17	T		\$null\$	0.000	77	92	62	5072.020	8377.800	1276.772	7100.828	False	Bronze	0.000
18	T		\$null\$	0.000	80	89	64	649.280	3669.400	2760.408	908.992	False	Silver	0.000
19	T		\$null\$	0.000	83	86	67	4745.250	8408.800	1843.450	6643.350	False	Bronze	0.000
20	T		\$null\$	0.000	70	86	53	9410.680	13715.800	540.848	13174.952	False	Bronze	0.000
21	T		0.000	0.000	82	87	64	1232.690	3761.800	2036.034	1725.766	False	Silver	0.000
22	T		\$null\$	1.000	83	90	68	-583.319	179.200	995.847	-816.647	True	Bronze	0.000
23	T		\$null\$	0.000	82	88	62	-1017.380	2584.800	1879.132	-1424.332	True	Bronze	0.000
24	T		\$null\$	0.000	82	91	63	2135.362	3221.400	2311.893	2989.507	False	Bronze	0.000
25	T		\$null\$	0.000	85	87	69	2495.593	3785.600	290.369	3495.231	False	Bronze	0.000
26	T		0.000	0.000	89	90	76	3247.940	5645.200	1699.084	4547.116	False	Bronze	0.000
27	T		\$null\$	0.000	95	90	71	1669.933	2545.200	179.294	2365.906	False	Bronze	0.000
28	T		\$null\$	0.000	80	91	60	7678.633	12661.600	1911.514	10750.066	False	Bronze	0.000
29	T		\$null\$	0.000	76	87	59	-115.560	2785.000	2926.184	-163.184	True	Silver	0.000
30	T		\$null\$	0.000	86	88	71	1861.520	5066.600	2460.472	2095.128	False	Silver	0.000
31	T		0.000	0.000	84	90	67	3052.200	4923.800	636.720	4287.080	False	Bronze	0.000
32	T		\$null\$	0.000	82	88	67	2957.13	9814.000	567.401	413.999	False	Bronze	0.000
33	T		\$null\$	1.000	78	89	63	1304.126	2629.200	803.424	1825.776	False	Bronze	0.000
34	T		\$null\$	1.000	83	92	70	807.450	5832.200	4501.176	1130.431	False	Silver	0.000
35	T		\$null\$	0.000	82	87	66	3986.500	6924.400	1343.300	5581.100	False	Bronze	0.000
36	T		0.000	0.000	82	87	67	8974.738	13924.400	1369.767	12854.633	False	Bronze	0.000
37	T		\$null\$	0.000	85	93	66	338.600	2130.800	1656.760	474.040	False	Bronze	0.000
38	T		\$null\$	0.000	74	88	62	7426.440	10441.200	44.184	10397.016	False	Bronze	0.000
39	T		\$null\$	0.000	87	92	73	144.410	1876.000	1673.826	202.174	False	Bronze	0.000
40	T		\$null\$	0.000	94	89	81	6849.030	9990.400	401.158	9588.642	False	Bronze	0.000
41	T		1.000	0.000	74	92	64	35.726	2412.200	2362.183	50.017	False	Silver	0.000
42	T		\$null\$	0.000	73	90	60	2953.820	5316.600	1183.532	4135.068	False	Bronze	0.000
43	T		\$null\$	0.000	87	92	69	507.470	1864.000	1146.942	710.458	False	Bronze	0.000

OK



unit 08 - IBM® SPSS® Modeler

Fields - 15 Records - 30000

Fields - 16 Records - 30000 Field Added - DIFFERENCE_CL_AS

Fields - 19 Records - 30000 Fields Added - CREDITLIMIT_DOLLAR AMOUNT_SPENT_DOLLAR DIFFERENCE_CL_AS_DOLLAR

ACME data part 1.dat DIFFERENCE_CL_AS _Dollar

Reclassify

Settings Annotations

Mode: Single Multiple

Reclassify info: New field Existing field

Reclassify field: GENDER

New field name: Reclassify1

Reclassify values:

Original value	New value
F	Female
M	Male
f	Female
m	Male

For unspecified values use: Original value Default value

OK Cancel Apply Reset

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Server: Local Server 123MB / 158MB

unit 08 - IBM® SPSS® Modeler

Fields - 15 Records - 30000

Fields - 16 Records - 30000 Field Added - DIFFERENCE_CL_AS

Fields - 19 Records - 30000 Fields Added - CREDITLIMIT_DOLLAR AMOUNT_SPENT_DOLLAR DIFFERENCE_CL_AS_DOLLAR

Fields - 20 Records - 30000 Field Added - SPENT_TOO MUCH

Fields - 21 Records - 30000 Field Added - SEGMENT

Fields - 22 Records - 30000 Field Added - BONUS

ACME data part 1.dat DIFFERENCE_CL_AS _Dollar SPENT_TOO MUCH SEGMENT

GENDER

Field: GENDER

Plot Appearance Output Annotations

Plot: Selected fields All flags (true values)

Field: GENDER

Overlay: Color: Normalize by color:

Sort: Alphabetic By count Proportional scale:

OK Run Cancel Apply Reset

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

Server: Local Server 123MB / 158MB

