

# IBM SPSS Statistics *Exercises*

**Yann Gouedo**

Data Scientist Leader – Machine Learning / Artificial Intelligence  
Marketing / Risk / Fraud / Maintenance  
IBM Certified Senior Data Scientist & IBM Certified Senior Architect



# IBM® SPSS® STATISTICS

## EXERCISE 1

Launch IBM®  
SPSS® Statistics

GUI Discovery

Upload  
Sales.sav

# IBM® SPSS® Statistics

## Software Launch

SPSS Statistics File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

Visible: 0 of 0 Variables

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

Welcome to IBM SPSS Statistics

IBM SPSS Statistics

New Files:

- New Dataset
- New Database Query...

Recent Files:

- .../POT SPSS Statistics EN/Sales.sav
- .../POT SPSS Statistics EN/Results.sav
- .../Insurances.sav
- Open another file...

What's New:

ROC Analysis

ROC Analysis supports the inference regarding a single AUC, precision-recall (PR) curves, and provides options for comparing two ROC curves that are generated from either independent groups or paired subjects.

ROC Curve

Sensitivity

1 - Specificity

Source of the Curve

Level of education

Predicted default, model 1

Visit the Community

for support and resources.

SPSS + dW

Getting Started:

Get Help and Support.

Get started with tutorials

Don't show this dialog in the future

Close

Data View Variable View

# IBM® SPSS® Statistics

## Sales.sav file Upload

SPSS Statistics File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Sales.sav [DataSet1] - IBM SPSS Statistics Data Editor

Visible: 12 of 12 Variables

	Order_Number	Client_Code	Product_Line_Number	Store_Code	Product_Code	Quantity	Sales_Price_AllTaxesIncl.	Prix_Vente_AllTaxesIncl.	Hour	Sales_Code	Amount	Transaction_Date	var	var	var	var	var	var
1	27343,000	126993,000	2,000	172,000	22029,000	1,000	26,000	13,800	1843,000	27,000	13,800	5-Feb-2014 00:00:00						
2	27343,000	126993,000	1,000	172,000	113353,000	1,000	2,400	1,270	1843,000	27,000	1,270	5-Feb-2014 00:00:00						
3	27343,000	126993,000	2,000	172,000	22029,000	1,000	26,000	13,800	1843,000	27,000	13,800	5-Feb-2014 00:00:00						
4	27343,000	126993,000	1,000	172,000	113353,000	1,000	2,400	1,270	1843,000	27,000	1,270	5-Feb-2014 00:00:00						
5	1488,000	151529,000	1,000	172,000	77410,000	1,000	21,870	21,870	1613,000	27,000	21,870	13-Jan-2014 00:00:00						
6	1488,000	151529,000	1,000	172,000	77410,000	1,000	21,870	21,870	1613,000	27,000	21,870	13-Jan-2014 00:00:00						
7	3277,000	214674,000	3,000	172,000	26338,000	1,000	52,800	42,240	1319,000	28,000	42,240	5-Jan-2015 00:00:00						
8	3277,000	214674,000	2,000	172,000	26342,000	1,000	58,800	47,040	1319,000	28,000	47,040	5-Jan-2015 00:00:00						
9	3277,000	214674,000	1,000	172,000	114063,000	1,000	26,220	16,780	1319,000	28,000	16,780	5-Jan-2015 00:00:00						
10	3277,000	214674,000	5,000	172,000	132740,000	1,000	6,540	5,230	1319,000	28,000	5,230	5-Jan-2015 00:00:00						
11	3277,000	214674,000	4,000	172,000	132971,000	1,000	5,200	4,160	1319,000	28,000	4,160	5-Jan-2015 00:00:00						
12	3277,000	214674,000	3,000	172,000	26338,000	1,000	52,800	42,240	1319,000	28,000	42,240	5-Jan-2015 00:00:00						
13	3277,000	214674,000	2,000	172,000	26342,000	1,000	58,800	47,040	1319,000	28,000	47,040	5-Jan-2015 00:00:00						
14	3277,000	214674,000	1,000	172,000	114063,000	1,000	26,220	16,780	1319,000	28,000	16,780	5-Jan-2015 00:00:00						
15	3277,000	214674,000	5,000	172,000	132740,000	1,000	6,540	5,230	1319,000	28,000	5,230	5-Jan-2015 00:00:00						
16	3277,000	214674,000	4,000	172,000	132971,000	1,000	5,200	4,160	1319,000	28,000	4,160	5-Jan-2015 00:00:00						
17	27356,000	318696,000	2,000	172,000	108006,000	1,000	11,870	11,870	1901,000	27,000	11,870	5-Feb-2014 00:00:00						
18	27356,000	318696,000	1,000	172,000	109942,000	1,000	15,870	15,870	1901,000	27,000	15,870	5-Feb-2014 00:00:00						
19	27356,000	318696,000	2,000	172,000	108006,000	1,000	11,870	11,870	1901,000	27,000	11,870	5-Feb-2014 00:00:00						
20	27356,000	318696,000	1,000	172,000	109942,000	1,000	15,870	15,870	1901,000	27,000	15,870	5-Feb-2014 00:00:00						
21	3455,000	318696,000	1,000	172,000	76107,000	1,000	61,870	48,540	1445,000	27,000	48,540	5-Feb-2015 00:00:00						
22	3455,000	318696,000	1,000	172,000	76107,000	1,000	61,870	48,540	1445,000	27,000	48,540	5-Feb-2015 00:00:00						
23	4596,000	438711,000	1,000	172,000	78270,000	1,000	30,270	30,270	1337,000	2,000	30,270	7-Jan-2014 00:00:00						
24	4596,000	438711,000	2,000	172,000	118557,000	1,000	44,800	44,800	1337,000	2,000	44,800	7-Jan-2014 00:00:00						
25	4602,000	438711,000	1,000	172,000	21887,000	1,000	27,070	13,740	1346,000	2,000	13,740	7-Jan-2014 00:00:00						
26	4596,000	438711,000	1,000	172,000	78270,000	1,000	30,270	30,270	1337,000	2,000	30,270	7-Jan-2014 00:00:00						
27	4596,000	438711,000	2,000	172,000	118557,000	1,000	44,800	44,800	1337,000	2,000	44,800	7-Jan-2014 00:00:00						
28	4602,000	438711,000	1,000	172,000	21887,000	1,000	27,070	13,740	1346,000	2,000	13,740	7-Jan-2014 00:00:00						
29	12085,000	445102,000	1,000	172,000	91903,000	1,000	56,000	56,000	1322,000	22,000	56,000	11-Jan-2013 00:00:00						
30	12085,000	445102,000	1,000	172,000	91903,000	1,000	56,000	56,000	1322,000	22,000	56,000	11-Jan-2013 00:00:00						
31	12026,000	495501,000	1,000	172,000	23768,000	1,000	32,400	32,400	1639,000	28,000	32,400	5-Jan-2013 00:00:00						
32	12026,000	495501,000	2,000	172,000	25441,000	1,000	41,340	41,340	1639,000	28,000	41,340	5-Jan-2013 00:00:00						
33	12026,000	495501,000	1,000	172,000	23768,000	1,000	32,400	32,400	1639,000	28,000	32,400	5-Jan-2013 00:00:00						

Data View Variable View

IBM SPSS Statistics Processor is ready

Unicode:ON

# IBM® SPSS® Statistics

## Variable View

SPSS Statistics File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Sales.sav [DataSet1] - IBM SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Order_Num...	Numeric	9	3		None	None	8	Right	Scale	Input
2	Client_Code	Numeric	9	3		None	None	8	Right	Scale	Input
3	Product_Lin...	Numeric	9	3		None	None	8	Right	Scale	Input
4	Store_Code	Numeric	9	3		None	None	8	Right	Scale	Input
5	Product_Code	Numeric	9	3		None	None	8	Right	Scale	Input
6	Quantity	Numeric	9	3		None	None	8	Right	Scale	Input
7	Sales_Price...	Numeric	9	3		None	None	8	Right	Scale	Input
8	Prix_Vente_...	Numeric	9	3		None	None	8	Right	Scale	Input
9	Hour	Numeric	9	3		None	None	8	Right	Scale	Input
10	Sales_Code	Numeric	9	3		None	None	8	Right	Scale	Input
11	Amount	Numeric	9	3		None	None	8	Right	Scale	Input
12	Transaction...	Date	20	0		None	None	18	Right	Scale	Input
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											
35											
36											

Data View Variable View

IBM SPSS Statistics Processor is ready

Unicode:ON

# IBM® SPSS® STATISTICS

## EXERCISE 2

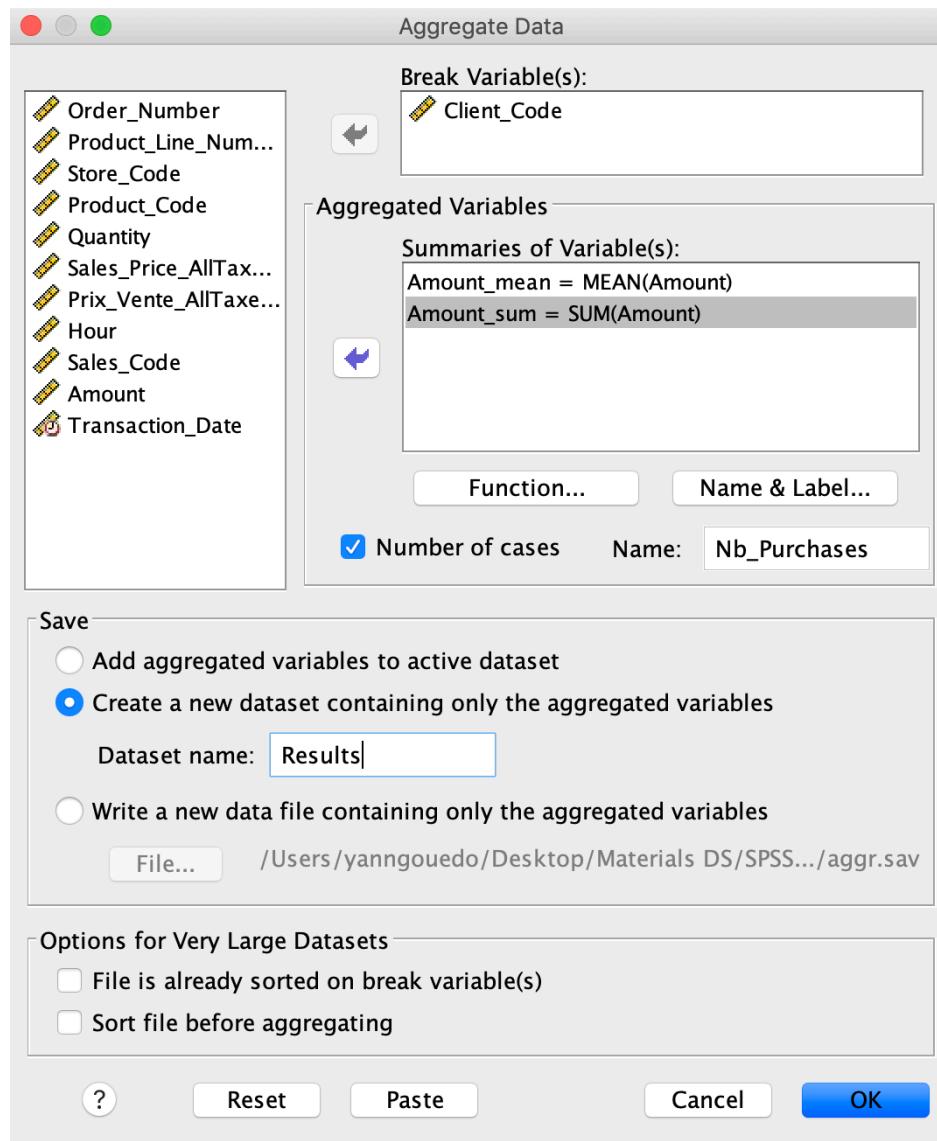
Aggregation of customer data to calculate the average basket, the amount of expenses and the number of orders per customer

Data export with the SPSS format

Visualize the syntax generated automatically

# IBM® SPSS® Statistics

## Data aggregation



# IBM® SPSS® Statistics

## Results Visualization

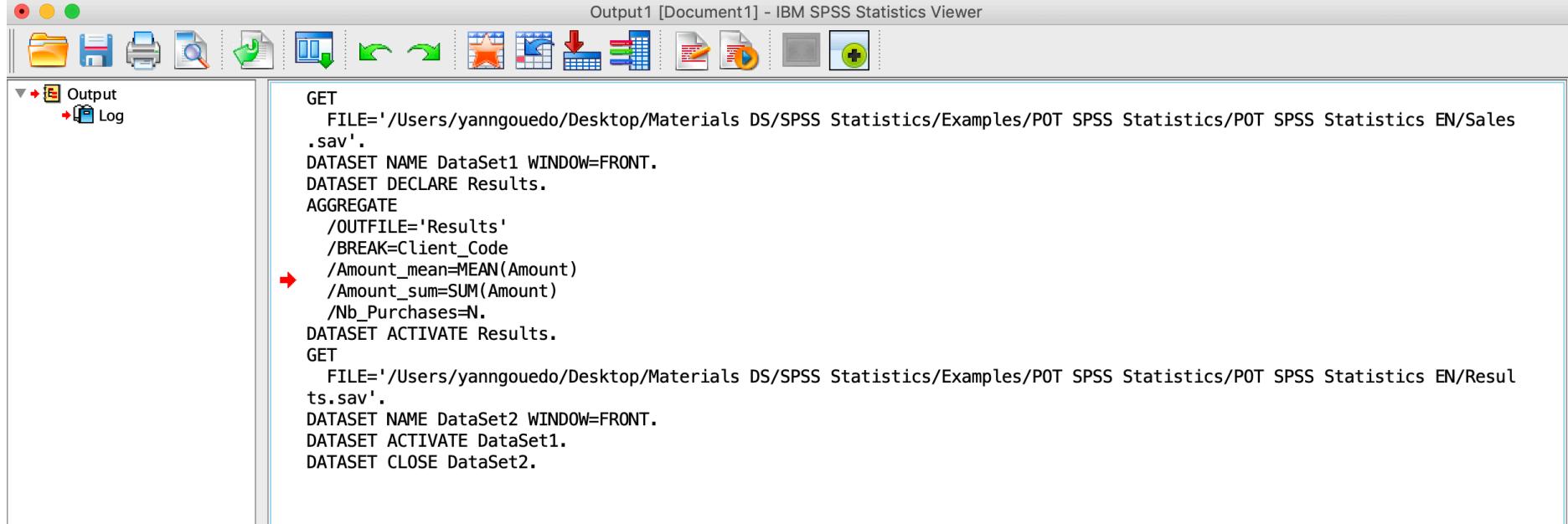
Untitled2 [Results] - IBM SPSS Statistics Data Editor

The screenshot shows the IBM SPSS Statistics Data Editor window titled "Untitled2 [Results]". The toolbar at the top includes icons for file operations (New, Open, Save, Print, etc.), data manipulation (Select Rows, Sort, Filter), and analysis (Descriptive Statistics, Crosstabs, etc.). The main area displays a table with 21 rows and 9 columns. The columns are labeled: Client\_Code, Amount\_mean, Amount\_sum, Nb\_Purchases, var, var, var, var, and var. The data consists of numerical values for each row. The "Data View" tab is selected at the bottom.

	Client_Code	Amount_mean	Amount_sum	Nb_Purchases	var	var	var	var
1	126993,000	7,54	30,14	4				
2	151529,000	21,87	43,74	2				
3	214674,000	23,09	230,90	10				
4	318696,000	25,43	152,56	6				
5	438711,000	29,60	177,62	6				
6	445102,000	56,00	112,00	2				
7	495501,000	25,74	205,88	8				
8	918238,000	67,34	134,68	2				
9	947752,000	10,67	21,34	2				
10	966788,000	58,67	352,02	6				
11	1090522,00	70,85	425,08	6				
12	1312145,00	78,67	157,34	2				
13	1315018,00	6,67	13,34	2				
14	1419162,00	11,28	22,56	2				
15	1419361,00	26,00	52,00	2				
16	1419451,00	97,03	194,06	2				
17	1419459,00	28,13	281,34	10				
18	1470727,00	56,40	112,80	2				
19	1470729,00	8,27	16,54	2				
20	1470754,00	57,67	230,68	4				
21	1470850,00	64,27	257,08	4				

# IBM® SPSS® Statistics

## Generated SPSS code Visualization



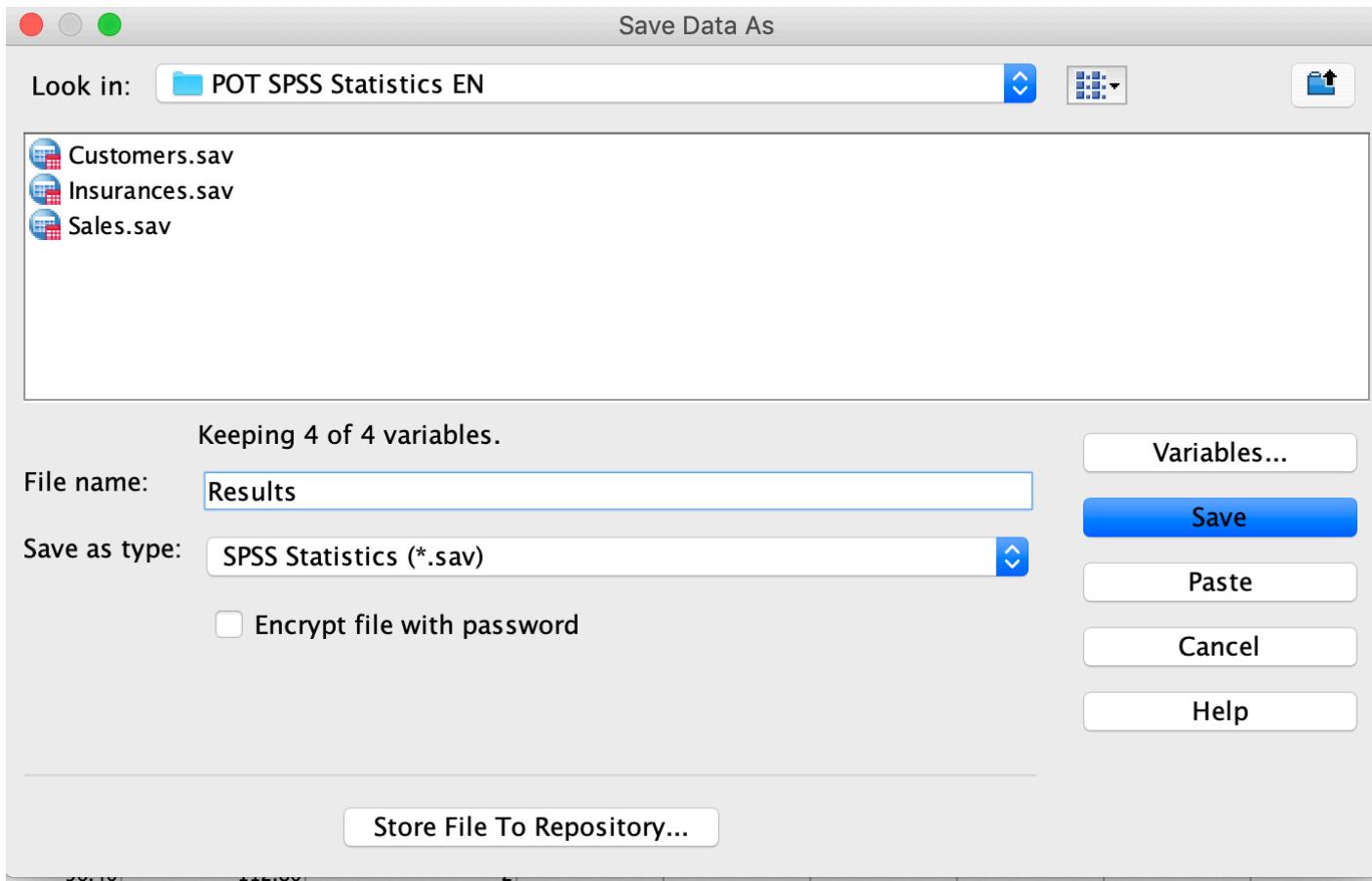
The screenshot shows the IBM SPSS Statistics Viewer window titled "Output1 [Document1] - IBM SPSS Statistics Viewer". The menu bar includes File, Edit, View, Analyze, Transform, Predictive Models, Utilities, Help, and a Language dropdown set to English (United States). The toolbar contains icons for opening files, saving, printing, and various data manipulation functions. On the left, a navigation pane shows a tree structure with "Output" selected and a "Log" entry. The main area displays the following SPSS syntax:

```
GET  
    FILE='/Users/yanngouedo/Desktop/Materials DS/SPSS Statistics/Examples/POT SPSS Statistics/POT SPSS Statistics EN/Sales.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.  
DATASET DECLARE Results.  
AGGREGATE  
    /OUTFILE='Results'  
    /BREAK=Client_Code  
    /Amount_mean=MEAN(Amount)  
    /Amount_sum=SUM(Amount)  
    /Nb_Purchases=N.  
DATASET ACTIVATE Results.  
GET  
    FILE='/Users/yanngouedo/Desktop/Materials DS/SPSS Statistics/Examples/POT SPSS Statistics/POT SPSS Statistics EN/Results.sav'.  
DATASET NAME DataSet2 WINDOW=FRONT.  
DATASET ACTIVATE DataSet1.  
DATASET CLOSE DataSet2.
```

A red arrow points to the line of code: `/Amount_mean=MEAN(Amount)`.

# IBM® SPSS® Statistics

## Data Exportation



# IBM® SPSS® STATISTICS

## EXERCISE 3

Customers.sav  
file upload

Merge between  
Results.sav and  
Customers.sav

Sorter by amount  
of cumulated  
expenses

Recodification of  
the Sex variable

Filter on the  
number of orders  
 $> 30$

Data Export

# IBM® SPSS® Statistics

## Customers.sav file Upload

Customers.sav [DataSet3] - IBM SPSS Statistics Data Editor

Visible: 7 of 7 Variables

	Client_Code	Adress1	PostalCode	City	Civility	LastName	FirstName	var	var	var
1	126993,000	151 RUE DE LA HAYE	62190,000	LILLERS	Mme	FAY	MONIQUE			
2	151529,000	10 RUE DE LA BRIE	77450,000	MONTRY	Mr	ABREMONT	FREDERIQUE			
3	214674,000	88 RUE DES POMMIERS	37300,000	JOUE LES TOURS	Mme	SADOSKY	STELLA			
4	318696,000	112 RUE ROBESPIERRE	59174,000	LA SENTINELLE	Mme	BOYER	Cecile			
5	438711,000	2273 ROUTE D AUDRUCIQ	62370,000	ST FOLQUIN	Mme	HEURTAUX	Liliane			
6	445102,000	FERME DE LA PLACE AUX PUITS	51530,000	ST MARTIN D ABOIS	Mr	RAUWEL	Jean			
7	495501,000	15 PLACE DES 2 CONILS	24100,000	BERGERAC	Mme	VALET	Maryse			
8	918238,000	5 RUE DES 3 DUGOIS	90300,000	VALDOIE	Mme	PINQUIER	KARINE			
9	947752,000	15 RUE DE LA PARFUMERIE	92600,000	ASNIERES SUR SEINE	Mr	BENOUMESSAD	Frederic			
10	966788,000	15 RUE DES 3 DUGOIS	67670,000	MOMMENHEIM	Mme	NGUYEN	LYDIE			
11	1090522,00	7 ROUTE DE CANTEPIE	76330,000	NORVILLE	Mme	GAUTHIER	Marie Josee			
12	1312145,00	6 RUE DU 5 FEVRIER	68230,000	WALBACH	Mme	PROTHIN	BERNADETTE			
13	1315018,00	31 RUE ERNEST THIERRY MIEG	90000,000	BELFORT	Mme	DESMONT	Vincent			
14	1419162,00	40 RUE DES CHARMES	86580,000	BIARD	Mme	GUILLAUT MEYER	Marie			
15	1419361,00	45 RUE GABRIEL PERI	76600,000	LE HAVRE	Mme	PONTAIS	Beilei			
16	1419451,00	327 AVENUE DU GENERAL LECLERC	6140,000	VENCE	Mme	JURY	Henriette			
17	1419459,00	6 RUE ANATOLE FRANCE	94270,000	LE KREMLIN BICETRE	Mme	PERSYN	Beatrice			
18	1470727,00	56 RUE PROUDHON	63000,000	CLERMONT FERRAND	Mme	ROUAULT	Kanle			
19	1470729,00	27 RUE DACOSSE	59495,000	LEFFRINCKOUCKE	Mme	MARSEILLE	Isabelle			
20	1470754,00	20 RUE DU SERGENT BOBILLOT	54000,000	NANCY	Mme	LOSSEL	Regine			
21	1470850,00	57 RUE D ALGER	62100,000	CALAIS	Mlle	THIBAULT	Marie Jose			
22	1470868,00	15 ROUTE DES COMBES	74400,000	CHAMONIX MONT BLANC	Mr	VERNY	Yvan			
23	1484775,00	LIEU DIT PAPON	3120,000	LAPALISSE	Mme	RICHARD	Simone			
24	1484998,00	7 RUE EMILE ALLIER	30000,000	NIMES	Mlle	BROSSET	Veronique			
25	1485003,00	53 RUE CONDORCET	59950,000	AUBY	Mme	TAMARELLE	Gwenaelle			
26	1542405,00	36 IMPASSE LES LILAS	80650,000	VIGNACOURT	Mme	SERGHERAERT	Charlotte			

Data View Variable View

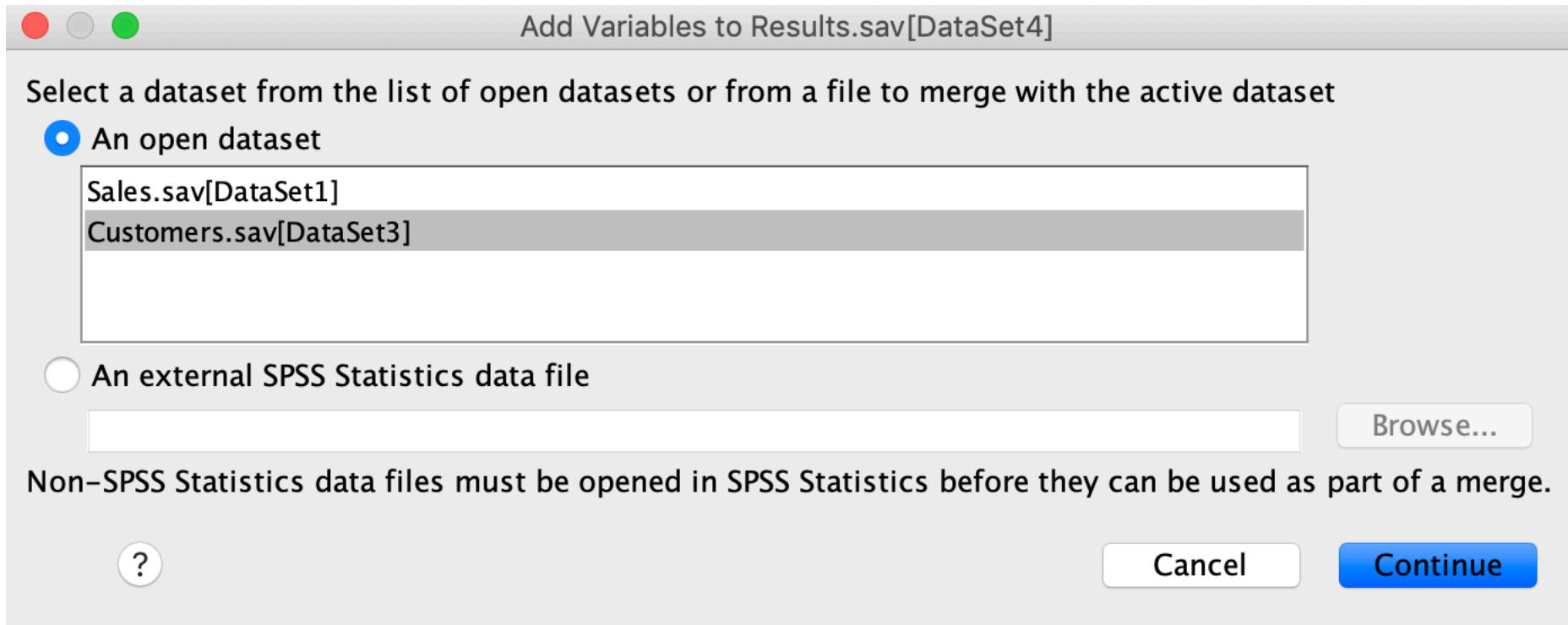
IBM SPSS Statistics Processor is ready

Unicode:ON

2019

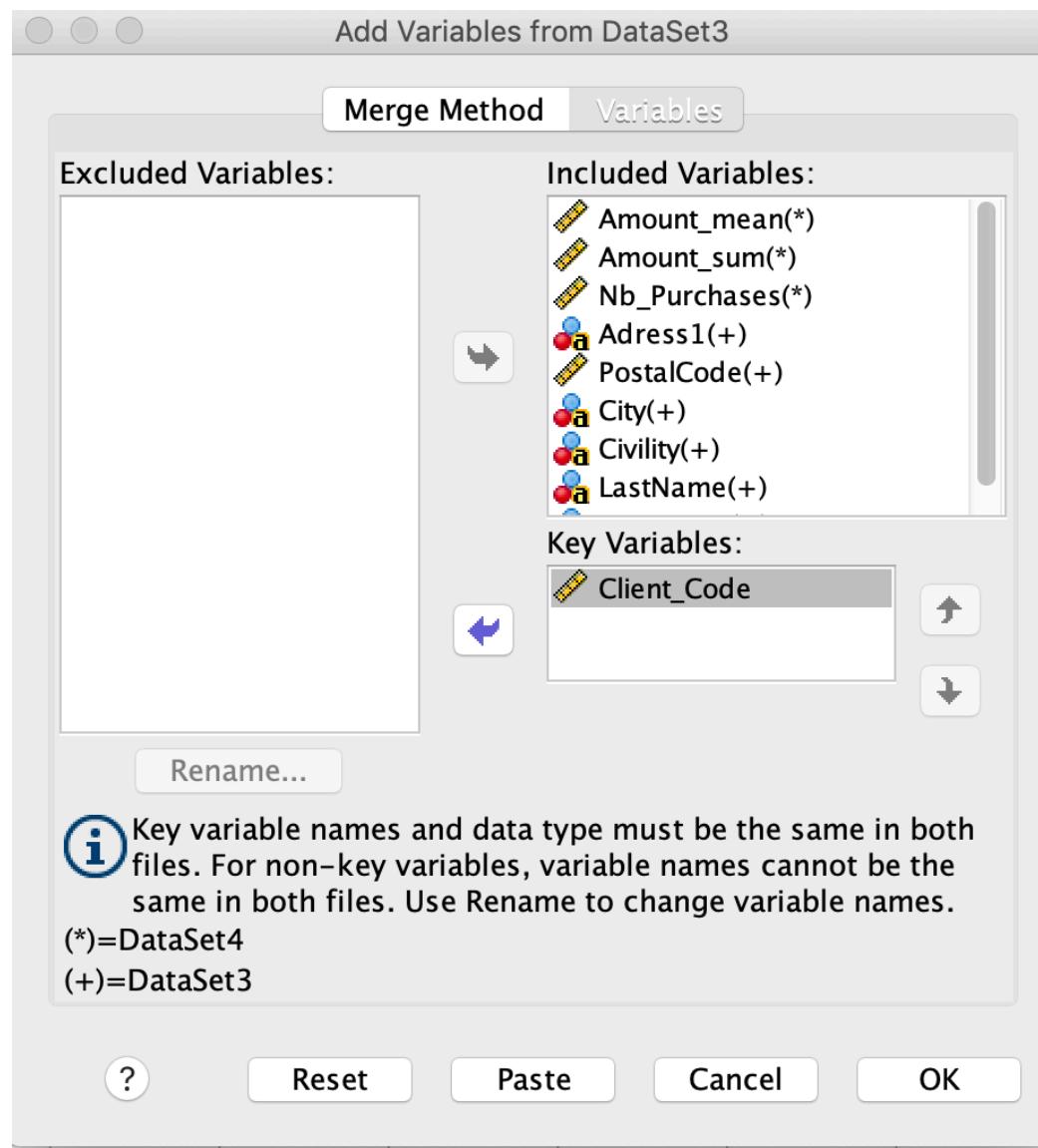
# IBM® SPSS® Statistics

## Merge between Results.sav and Customers.sav



# IBM® SPSS® Statistics

## Data Merging



# IBM® SPSS® Statistics Result

Results.sav [DataSet7] - IBM SPSS Statistics Data Editor

Visible: 10 of 10 Variables

	Client_Code	Amount_mean	Amount_sum	Nb_Purchases	Adress1	PostalCode	City	Civility	LastName	FirstName	var	var
1	.	.	.	.		83110,000	SANARY	\$null\$	\$null\$	\$null\$		
2	126993,000	7,54	30,14	4	151 RUE DE LA HAYE	62190,000	LILLERS	Mme	FAY	MONIQUE		
3	151529,000	21,87	43,74	2	10 RUE DE LA BRIE	77450,000	MONTRY	Mr	ABREMONT	FREDERIQUE		
4	214674,000	23,09	230,90	10	88 RUE DES POMMIERS	37300,000	JOUE LES TOURS	Mme	SADOSKY	STELLA		
5	318696,000	25,43	152,56	6	112 RUE ROBESPIERRE	59174,000	LA SENTINELLE	Mme	BOYER	Cecile		
6	438711,000	29,60	177,62	6	2273 ROUTE D AUDRUICQ	62370,000	ST FOLQUIN	Mme	HEURTAUX	Liliane		
7	445102,000	56,00	112,00	2	FERME DE LA PLACE AUX PUITS	51530,000	ST MARTIN D ABLAIS	Mr	RAUWEL	Jean		
8	495501,000	25,74	205,88	8	15 PLACE DES 2 CONILS	24100,000	BERGERAC	Mme	VALET	Maryse		
9	918238,000	67,34	134,68	2	5 RUE DES 3 DUGOIS	90300,000	VALDOIE	Mme	PINQUIER	KARINE		
10	947752,000	10,67	21,34	2	15 RUE DE LA PARFUMERIE	92600,000	ASNIERES SUR SEINE	Mr	BENOUMESSAD	Frederic		
11	966788,000	58,67	352,02	6	15 RUE DES 3 DUGOIS	67670,000	MOMMENHEIM	Mme	NGUYEN	LYDIE		
12	1090522,00	70,85	425,08	6	7 ROUTE DE CANTEPIE	76330,000	NORVILLE	Mme	GAUTHIER	Marie Josee		
13	1312145,00	78,67	157,34	2	6 RUE DU 5 FEVRIER	68230,000	WALBACH	Mme	PROTHIN	BERNADETTE		
14	1315018,00	6,67	13,34	2	31 RUE ERNEST THIERRY MIEG	90000,000	BELFORT	Mme	DESMONT	Vincent		
15	1419162,00	11,28	22,56	2	40 RUE DES CHARMES	86580,000	BIARD	Mme	GUILLAUT MEYER	Marie		
16	1419361,00	26,00	52,00	2	45 RUE GABRIEL PERI	76600,000	LE HAVRE	Mme	PONTAIS	Beilei		
17	1419451,00	97,03	194,06	2	327 AVENUE DU GENERAL LECLERC	6140,000	VENCE	Mme	JURY	Henriette		
18	1419459,00	28,13	281,34	10	6 RUE ANATOLE FRANCE	94270,000	LE KREMLIN BICETRE	Mme	PERSYN	Beatrice		
19	1470727,00	56,40	112,80	2	56 RUE PROUDHON	63000,000	CLERMONT FERRAND	Mme	ROUAULT	Karie		
20	1470729,00	8,27	16,54	2	27 RUE DACOSSE	59495,000	LEFFRINCKOUCKE	Mme	MARSEILLE	Isabelle		
21	1470754,00	57,67	230,68	4	20 RUE DU SERGENT BOBILLOT	54000,000	NANCY	Mme	LOSSEL	Regine		
22	1470850,00	64,27	257,08	4	57 RUE D ALGER	62100,000	CALAIS	Mlle	THIBAULT	Marie Jose		
23	1470868,00	5,47	10,94	2	15 ROUTE DES COMBES	74400,000	CHAMONIX MONT BLANC	Mr	VERNY	Yvan		
24	1484775,00	40,00	80,00	2	LIEU DIT PAPON	3120,000	LAPALISSE	Mme	RICHARD	Simone		
25	1484998,00	55,23	220,92	4	7 RUE EMILE ALLIER	30000,000	NIMES	Mlle	BROSSET	Veronique		
26	1485003,00	7,87	31,48	4	53 RUE CONDORCET	59950,000	AUBY	Mme	TAMARELLE	Gwenaelle		
27	1542405,00	76,31	1678,92	22	36 IMPASSE LES LILAS	80650,000	VIGNACOURT	Mme	SERGHERAERT	Charlotte		
28	1556672,00	14,27	57,08	4	13 LOTISSEMENT CHAMP MONSIEUR	55100,000	CHARNY SUR MEUSE	Mme	GOULU	Madeleine		
29	1560817,00	47,58	95,16	2	15 RUE DE DASLE	25400,000	AUDINCOURT	Mme	TINGAUD	SANDRINE		
30	1568148,00	20,27	40,54	2	29D ROUTE DES AGRIONS	7130,000	SOYONS	Mme	CHALLON	Madeleine		
--	1600043,00	57,21	114,68	2		52370,000	ONIAV	Mme	EDEV MEDINA	Stephanie		

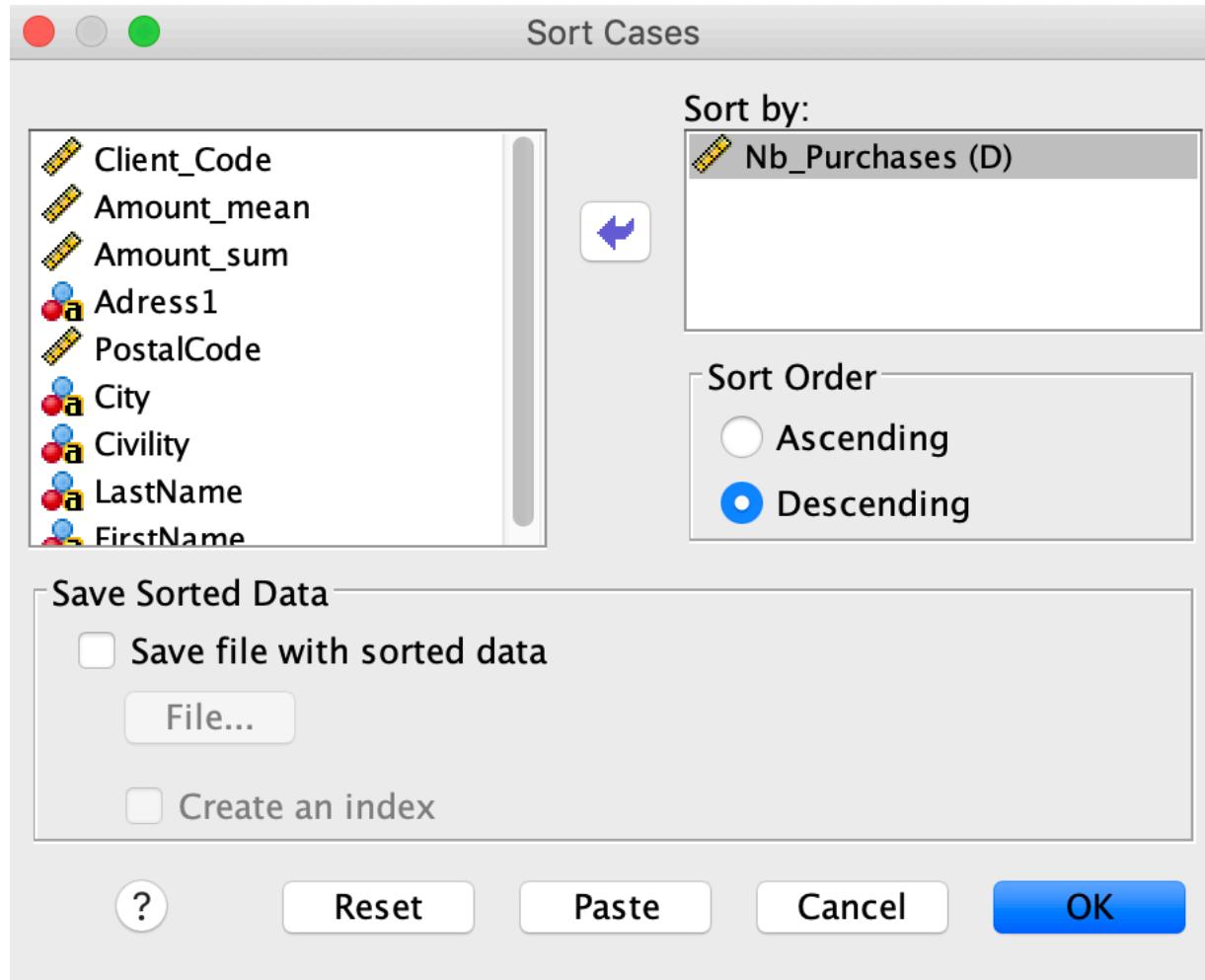
Data View Variable View

IBM SPSS Statistics Processor is ready

Unicode:ON

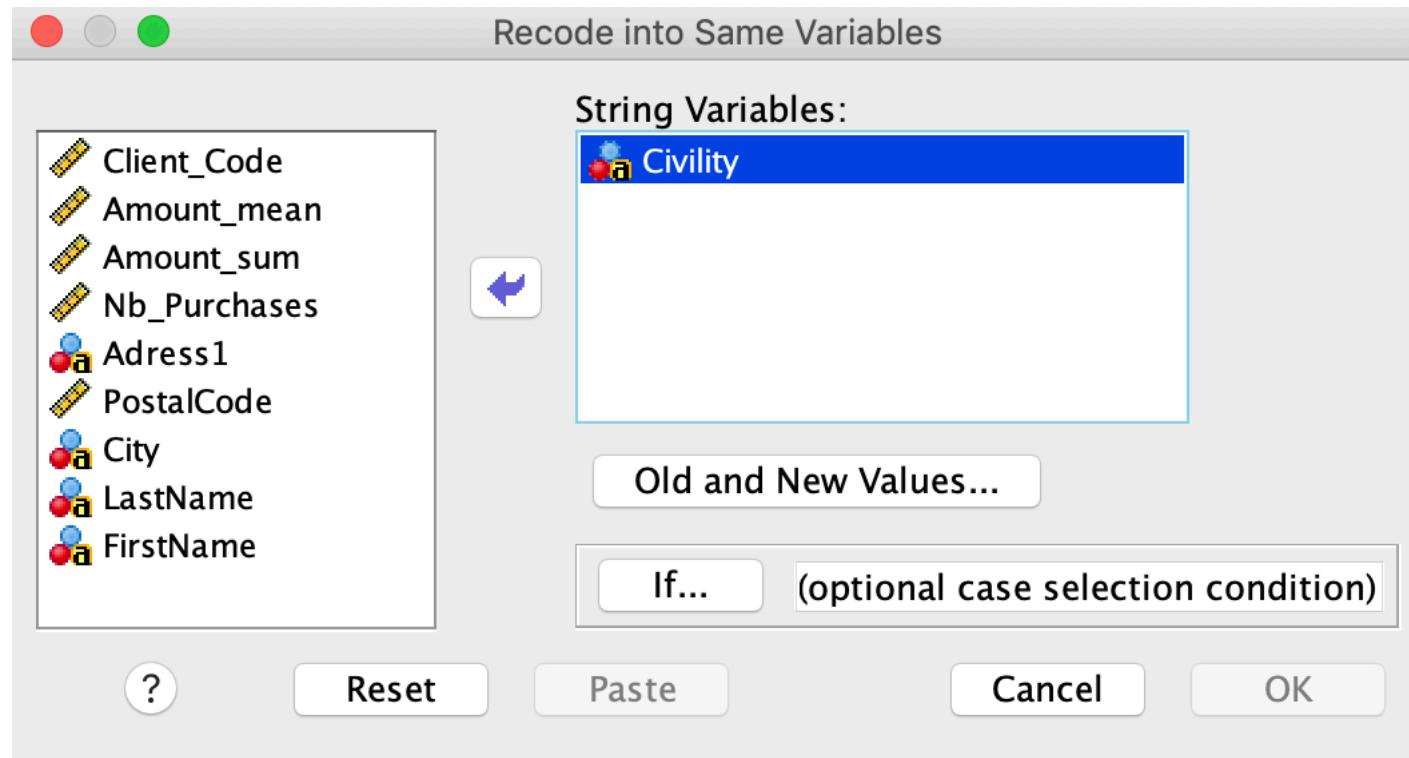
# IBM® SPSS® Statistics

## Data Sorting



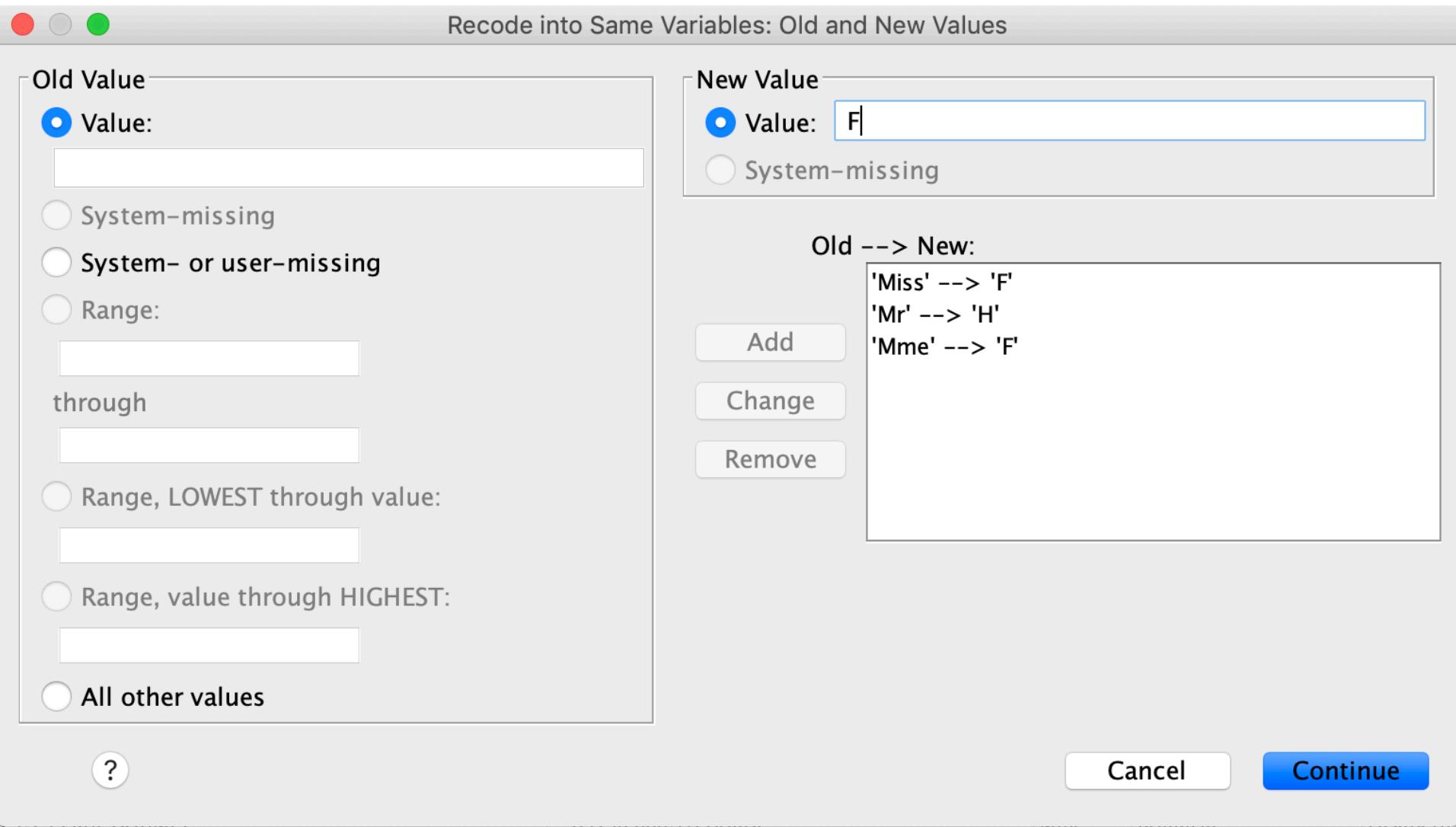
# IBM® SPSS® Statistics

## Recodification



# IBM® SPSS® Statistics

## Recodification



# IBM® SPSS® Statistics

## Recodification

Results.sav [DataSet1] - IBM SPSS Statistics Data Editor

Visible: 10 of 10 Variables

	Client_Code	Amount_mean	Amount_sum	Nb_Purchases	Adress1	PostalCode	City	Civility	LastName	FirstName	var	var
1	4764712,00	16,88	978,82	58	8 RUE MARCEL PAGNOL	86530,000	CENON SUR VIENNE	F	BILLON	Jessica		
2	5721627,00	5,85	280,62	48	8 RUE DE BOUCHAVESNES	80200,000	ALLAINES	F	JANY	MARIE-PIERRE		
3	2335946,00	11,94	525,36	44		83210,000	BELGENTIER	H	LEDUC	Gregory		
4	4399574,00	8,88	372,96	42	19 AVENUE ANNE ET ALBERT PROVOST	59910,000	BONDUES	F	HUBERT	Claire		
5	2320710,00	16,73	602,40	36		83400,000	HYERES	F	ANOMAN	DENISE		
6	2327419,00	28,67	974,74	34	774 AV DE LA MALOGINESTE	83140,000	SIX FOURES LES PLAGES	F	VIOLI	Dalila		
7	5178061,00	4,10	139,56	34	48 RUE DE LA CERISAIE	38360,000	SASSENAGE	F	JACQUES	Jeannine		
8	2304797,00	15,79	505,32	32	ALLEE DES TAMARIS	83550,000	VIDAUBAN	H	HOCDEE	Aurelien		
9	2793006,00	45,32	1450,20	32	84 vbis Avenue de Fontaine bleau	94270,000	LE KREMLIN BICETRE	F	GOASDOUE	Clotilde		
10	2837223,00	32,12	1027,78	32	36 RUE SAINT CHRISTOPHE	2200,000	SOISSONS	F	PARFOND	Delphine		
11	2273534,00	37,18	1115,30	30	98 AVENUE DE BIR HAKEIM	83110,000	SANARY SUR MER	F	FUMET	Stephanie		
12	6889872,00	20,40	612,14	30	1761 ROUTE DE PERNES	84210,000	ALTHEN DES PALUDS	F	DESCA	EMMANUELLE		
13	2954382,00	73,11	2047,18	28	17 ROUTE DE MARCHON	1100,000	OYONNAX	F	SABLON	Marie Pierre		
14	4575407,00	12,64	353,90	28	26 RUE DES TOURNANFIS	94360,000	BRY SUR MARNE	F	RAISON	PASCAL		
15	5987877,00	14,22	398,04	28	4 RUE JACQUES BREL	14440,000	DOUVRES LA DELIVRANTE	F	BRUNA	Clarissa		
16	2332180,00	19,06	495,66	26		83320,000	CARCAIRANE	F	ROUDON	YVETTE		
17	2458199,00	63,45	1649,68	26	LA VETTE	241,000	LE FLEIX	F	GARREC	Laurence		
18	4109904,00	3,90	101,38	26	47 RUE de FONTAINE	21000,000	DIJON	H	VERAZZI	GREGORY		
19	7486493,00	5,06	131,68	26	212 avenue delattre de tassigny	59700,000	marcq en baroeul	F	POTHIER	Monique		
20	2273933,00	19,98	479,42	24	LIEU DIT LA BOURGIAT	42570,000	ST HEAND	F	DERRADJI	Jacqueline		
21	2274103,00	5,36	128,72	24	8 ROUTE DE LA COTE VALLEE	76133,000	ST MARTIN DU BEC	F	AMSELLEM	Bernard		
22	2275650,00	24,27	582,46	24	LA FARIGOULETTE IMP FLORY	83160,000	LA VALETTE DU VAR	F	BOUSQUET	Nathalie		
23	2307972,00	45,24	1085,68	24	243 RUE DOUMET	83130,000	LA GARDE	F	BARRIERE	Elisabeth		
24	2324777,00	,00	,00	24		83160,000	LA VALETTE	F	BEAUVAIS	Jehanne		
25	2968138,00	17,06	409,36	24	5 RUE DES PERDRIX	91490,000	MOIGNY-SUR-ECOLE	F	DIKI	CORINE		
26	3119824,00	37,98	911,42	24	22 ROUTE DE BROCHARD	33370,000	LOUPES	F	EL KENZ	DENISE		
27	3192856,00	2,49	59,86	24	12 RUE ERIK SATIE	24100,000	BERGERAC	F	CONFINO	Fanny		
28	3227714,00	9,87	236,88	24	8 RUE DU HAMEAU	34430,000	ST JEAN DE VEDAS	H	CHAMPAGNE	NOUR		
29	1542405,00	76,31	1678,92	22	36 IMPASSE LES LILAS	80650,000	VIGNACOURT	F	SERGHERAERT	Charlotte		
30	1791398,00	28,77	632,98	22	3 IMPASSE POUSSIN	95120,000	ERMONT	F	PIRAS	Paula		
31	2248419,00	8,34	183,44	22	1410 RUE DE L ECLUSE CARREE	62730,000	LES ATTAQUES	F	PAULHAC	Christine		

Data View Variable View

IBM SPSS Statistics Processor is ready

Unicode:ON

# IBM® SPSS® Statistics Filtering

Select Cases: If

The dialog box shows a list of variables on the left, with 'Nb\_Purchases' selected. The main area displays the filter condition 'Nb\_Purchases > 30'. A numeric keypad is provided for entering values, and a function group list and functions list are on the right.

Nb\_Purchases > 30

Function group:

- All
- Arithmetic
- CDF & Noncentral CDF
- Conversion
- Current Date/Time
- Date Arithmetic
- Date Creation

Functions and Special Variables:

Cancel Continue

# IBM® SPSS® Statistics

## Result

Results.sav [DataSet1] - IBM SPSS Statistics Data Editor

Visible: 11 of 11 Variables

	Client_Code	Amount_mean	Amount_sum	Nb_Purchases	Adress1	PostalCode	City	Civility	LastName	FirstName	filter_\$	var
1	4764712,00	16,88	978,82		58 8 RUE MARCEL PAGNOL	86530,000	CENON SUR VIENNE	F	BILLON	Jessica		1
2	5721627,00	5,85	280,62		48 8 RUE DE BOUCHAVESNES	80200,000	ALLAINES	F	JANY	MARIE-PIERRE		1
3	2335946,00	11,94	525,36		44	83210,000	BELGENTIER	H	LEDUC	Gregory		1
4	4399574,00	8,88	372,96		42 19 AVENUE ANNE ET ALBERT PROUVOST	59910,000	BONDUES	F	HUBERT	Claire		1
5	2320710,00	16,73	602,40		36	83400,000	HYERES	F	ANOMAN	DENISE		1
6	2327419,00	28,67	974,74		34 774 AV DE LA MALOGINESTE	83140,000	SIX FOURES LES PLAGES	F	VIOLI	Dalila		1
7	5178061,00	4,10	139,56		34 48 RUE DE LA CERISAIE	38360,000	SASSENAGE	F	JACQUES	Jeannine		1
8	2304797,00	15,79	505,32		32 ALLEE DES TAMARIS	83550,000	VIDAUBAN	H	HOCDEE	Aurelien		1
9	2793006,00	45,32	1450,20		32 84 vbis Avenue de Fontaine bleau	94270,000	LE KREMLIN BICETRE	F	GOASDOUÉ	Clotilde		1
10	2837223,00	32,12	1027,78		32 36 RUE SAINT CHRISTOPHE	2200,000	SOISSONS	F	PARFOND	Delphine		1
11	2273534,00	37,18	1115,30		30 98 AVENUE DE BIR HAKEIM	83110,000	SANARY SUR MER	F	FUMET	Stephanie		0
12	6889872,00	20,40	612,14		30 1761 ROUTE DE PERNES	84210,000	ALTHEN DES PALUDS	F	DESCA	EMMANUELLE		0
13	2954382,00	73,11	2047,18		28 17 ROUTE DE MARCHON	1100,000	OYONNAX	F	SABLON	Marie Pierre		0
14	4575407,00	12,64	353,90		28 26 RUE DES TOURNANFIS	94360,000	BRY SUR MARNE	F	RAISON	PASCAL		0
15	5987877,00	14,22	398,04		28 4 RUE JACQUES BREL	14440,000	DOUVRES LA DELIVRANDE	F	BRUNA	Clarissa		0
16	2332180,00	19,06	495,66		26	83320,000	CARCAIRANE	F	ROUDON	YVETTE		0
17	2458199,00	63,45	1649,68		26 LA VETTE	241,000	LE FLEIX	F	GARREC	Laurence		0
18	4109904,00	3,90	101,38		26 47 RUE de FONTAINE	21000,000	DIJON	H	VERAZZI	GREGORY		0
19	7486493,00	5,06	131,68		26 212 avenue delattre de tassigny	59700,000	marcq en baroeul	F	POTHIER	Monique		0
20	2273933,00	19,98	479,42		24 LIEU DIT LA BOURGIAT	42570,000	ST HEAND	F	DERRADJI	Jacqueline		0
21	2274103,00	5,36	128,72		24 8 ROUTE DE LA COTE VALLEE	76133,000	ST MARTIN DU BEC	F	AMSELLEM	Bernard		0
22	2275650,00	24,27	582,46		24 LA FARIGOULETTE IMP FLORY	83160,000	LA VALETTE DU VAR	F	BOUSQUET	Nathalie		0
23	2307972,00	45,24	1085,68		24 243 RUE DOUMET	83130,000	LA GARDE	F	BARRIERE	Elisabeth		0
24	2324777,00	,00	,00		24	83160,000	LA VALETTE	F	BEAUVAIS	Jehanne		0
25	2968138,00	17,06	409,36		24 5 RUE DES PERDRIX	91490,000	MOIGNY-SUR-ECOLE	F	DIKI	CORINE		0
26	3119824,00	37,98	911,42		24 22 ROUTE DE BROCHARD	33370,000	LOUPES	F	EL KENZ	DENISE		0
27	3192856,00	2,49	59,86		24 12 RUE ERIK SATIE	24100,000	BERGERAC	F	CONFINO	Fanny		0
28	3227714,00	9,87	236,88		24 8 RUE DU HAMEAU	34430,000	ST JEAN DE VEDAS	H	CHAMPAGNE	NOUR		0
29	1542405,00	76,31	1678,92		22 36 IMPASSE LES LILAS	80650,000	VIGNACOURT	F	SERGHERAERT	Charlotte		0
30	1791398,00	28,77	632,98		22 3 IMPASSE POUSSIN	95120,000	ERMONT	F	PIRAS	Paula		0
31	2248419,00	8,34	183,44		22 1410 RUE DE L'ECLUSE CARREE	62730,000	LES ATTAQUES	F	PAULHAC	Christine		0

Data View Variable View

IBM SPSS Statistics Processor is ready

Unicode:ON Filter On

# IBM® SPSS® STATISTICS

## EXERCISE 4

Insurances.sav  
file Upload

Descriptive  
statistics

Crossed table  
(type OLAP)

Results export  
(HTML format)

# IBM® SPSS® Statistics

## Data

Insurances.sav [DataSet1] - IBM SPSS Statistics Data Editor

Visible: 16 of 16 Variables

	Customer_Id	Brand	Usage	Socio_Professional_Category	Power	Value	Pricing_Zone	Marital_Status	Year_Driving_License	Garage_Mode	Age	Vehicle_Oldness
1	1000758	Peugeot	Private & Professional	Workers (W)	6	6	3	Married	1985	No	53	16
2	1000980	Citroën	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	8	8	1	Married	1990	Yes	41	13
3	1001134	Peugeot	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	8	7	9	Married	2000	Yes	30	8
4	1001267	Renault	Private & Work journey	Employees (E)	9	8	8	Married	1981	Yes	53	11
5	1003548	Renault	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	6	5	5	Married	1995	Yes	36	5
6	1003584	Renault	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	3	2	5	Single	1997	No	35	9
7	1003585	Saab	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	7	8	5	Single	1996	No	34	8
8	1004401	Chrysler	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	6	9	4	Unknown	.	Yes	.	19
9	1005877	Renault	Private & Professional	Intermediate professions (IP)	9	12	5	Married	1982	Yes	50	13
10	1006037	Mercedes-Benz	Private	Intermediate professions (IP)	5	5	4	Single	2006	Yes	25	8
11	1006384	Renault	Private & Work journey	Employees (E)	10	10	3	Married	1983	Yes	50	10
12	1006675	Saab	Private & Work journey	Other people without professional activity (OPWPA)	8	9	8	Single	2005	No	28	13
13	1007010	Renault	Private & Work journey	Craftsmen, Shopkeepers and Businessmen (CSB)	10	14	7	Unknown	.	Yes	.	4
14	1007042	Peugeot	Private	Intermediate professions (IP)	6	5	9	Single	2002	No	26	11
15	1007538	Renault	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	7	6	3	Married	1987	Yes	44	4
16	1008094	Renault	Private & Work journey	Employees (E)	7	7	2	Married	1967	No	76	9
17	1008889	Alfa Romeo	Private	Senior Managers and Professional Occupations (SMP...)	8	11	7	Married	1995	Yes	.	15
18	1009969	Volkswagen	Private & Work journey	Employees (E)	6	7	7	Married	1989	Yes	44	10
19	1010805	Audi	Private & Professional	Farmers (F)	11	13	3	Single	1996	Yes	37	11
20	1011118	Peugeot	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	8	7	3	Married	2000	No	31	5
21	1011717	Audi	Private	Senior Managers and Professional Occupations (SMP...)	8	12	5	Unknown	1991	No	42	17
22	1012379	Ford	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	5	4	8	Married	1992	Yes	39	10
23	1012885	Peugeot	Private & Work journey	Employees (E)	5	4	6	Married	1989	No	44	19
24	1012975	Peugeot	Private	Senior Managers and Professional Occupations (SMP...)	6	5	5	Married	1999	No	37	10
25	1013306	Renault	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	3	3	5	Married	1966	Yes	64	12
26	1013347	Renault	Private & Work journey	Employees (E)	9	8	3	Unknown	2002	Yes	43	11
27	1013757	Saab	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	5	5	5	Unknown	1997	No	36	6
28	1014149	Peugeot	Private & Work journey	Other people without professional activity (OPWPA)	4	3	3	Single	2005	Yes	25	8
29	1014350	Fiat	Private & Work journey	Employees (E)	6	8	5	Married	1985	Yes	51	16
30	1014634	Renault	Private & Work journey	Employees (E)	8	7	2	Married	1983	Yes	47	10
31	1014982	Renault	Private & Work journey	Senior Managers and Professional Occupations (SMP...)	7	8	3	Married	1987	Yes	45	15

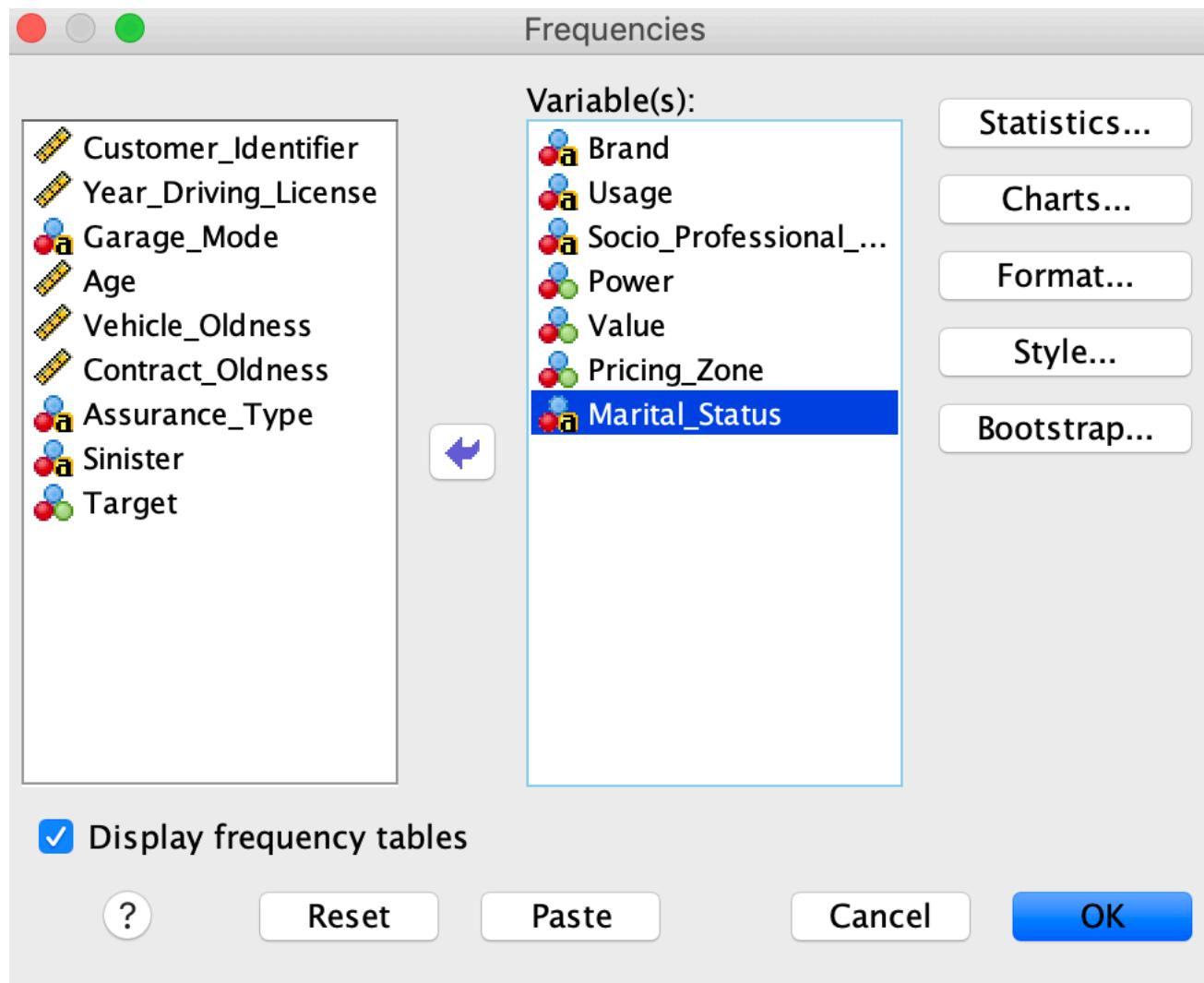
Data View Variable View

IBM SPSS Statistics Processor is ready

Unicode:ON

# IBM® SPSS® Statistics

## Descriptive Statistics



# IBM® SPSS® Statistics Results

Output2 [Document2] - IBM SPSS Statistics Viewer

Output Log Frequencies Frequency Table Log Frequencies Frequency Table Bar Chart

Marital\_Status

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Divorced	25	1,2	1,2
	Married	1373	63,6	63,6
	Single	431	20,0	84,7
	Unknown	331	15,3	100,0
Total		2160	100,0	

Bar Chart

Brand

Brand	Frequency
Alfa Romeo	~20
Audi	~30
Autres	~10
BMW	~10
Chrysler	~50
Citroën	~350
Fiat	~100
Ford	~150
Honda	~5
Kia	~5
Lancia	~5
Mazda	~5
Mercedes	~5
Nissan	~5
Porsche	~400
Renault	~750
Saab	~80
Smart	~20
Toyota	~40
Volkswagen	~70
Volvo	~20

IBM SPSS Statistics Processor is ready Unicode:ON

# IBM® SPSS® Statistics

## Cross tables

Custom Tables

Table    Titles    Test Statistics    Options

Variables:

- Customer\_Id...
- Brand
- Usage
- Socio\_Profess...
- Power
- Value
- Pricing\_Zone
- Marital\_Status
- Year\_Driving\_...
- Garage\_Mode
- Age
- Vehicle\_Oldn...
- Contract\_Old...
- Assurance\_Ty...
- Sinister

Columns

				Brand	
				Category 1	Category 2
				Count	Count
Socio_Professional_Category	Category 1	Usage	Category 1	nnnn	nnnn
		Category 2	nnnn	nnnn	
	Category 2	Usage	Category 1	nnnn	nnnn
		Category 2	nnnn	nnnn	

Rows

Categories:

- Category 1
- Category 2

Define

N% Summary Statistics...  
Categories and Totals...

Summary Statistics

Position: Columns  Hide

Source: Row Variables

Category Position:  
Default

?

Reset

Paste

Cancel

OK

# IBM® SPSS® Statistics Results

Output2 [Document2] - IBM SPSS Statistics Viewer

The screenshot shows the IBM SPSS Statistics Viewer interface. On the left, there's a navigation pane with a tree view of output documents, tables, and log files. A red arrow points from the 'Custom Tables' section in the tree to the table below. Another red arrow points from the 'Table 1' entry in the tree to the bottom of the table.

**Marital\_Status**

Marital_Status	Count
Divorced	1
Married	19
Single	1
Unknown	1

\* Custom Tables.  
CTABLES  
/VLABELS VARIABLES=Socio\_Professional\_Category Usage Brand DISPLAY=LABEL  
/TABLE Socio\_Professional\_Category > Usage [COUNT F40.0] BY Brand  
/CATEGORIES VARIABLES=Socio\_Professional\_Category Usage Brand ORDER=A KEY=VALUE EMPTY=EXCLUDE  
/CRITERIA CILEVEL=95.

**Custom Tables**

Socio_Professional_Cate gory	Usage	Private & Professional	Brand								
			Alfa Romeo Count	Audi Count	Autres Count	BMW Count	Chrysler Count	Citroën Count	Fiat Count	Ford Count	Hc Count
Craftsmen, Shopkeepers and Businessmen (CSB)	Private & Work journey	2	0	0	0	5	17	3	3		
	Employees (E)	1	0	0	0	0	5	3	2		
Employees (E)	Private	0	1	0	0	1	6	2	5		
	Private & Professional	0	0	0	0	0	3	1	3		
Farmers (F)	Private & Work journey	4	3	2	4	3	39	14	16		
	Private & Professional	2	2	0	1	3	25	2	6		
Intermediate professions (IP)	Private	3	2	2	1	2	28	8	8		
	Private & Professional	0	0	0	0	2	9	0	2		
Other people without professional activity (OPWPA)	Private & Work journey	0	0	0	0	0	0	0	0		
	Private	0	0	0	0	1	0	0	0		
Retirees (R)	Private & Professional	1	0	0	0	0	0	0	0		
	Private	5	0	0	2	9	33	10	9		
Senior Managers and Professional Occupations (SMPO)	Private	2	2	3	3	6	17	9	15		
	Private & Professional	3	1	0	0	3	26	8	17		
Workers (W)	Private & Work journey	5	4	1	4	13	103	37	54		
	Private & Professional	2	0	0	0	2	12	2	7		
	Private & Work journey	0	0	0	0	0	1	0	0		

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 473, W: 2078 pt.

# IBM® SPSS® Statistics

## Table looks options

TableLook: Default

TableLook Files:

- <As Displayed>
- <System Default>
- <Classic Default>
- Easter Egg
- Cobalt
- Simple
- Soft
- AnalyticsPlatform
- MonocolorContrast
- ContrastAlternate
- WebReportBlue
- Academic
- CompactBoxed

Reset all cell formats to the TableLook

Sample

Table Title<sup>a,b</sup>

Layer:layer1

		bbbb1	aaaa	bbbb2	aaaa	
ddd	cccc	aaa1	aaaa2	aaa1	aaaa2	
dddd1	cccc1	0	abcd	212.4	abcd	
	cccc2	88.6	abcd	83.65	abcd	
group	dddd2	cccc1	105	abcd	58.53	abcd
	cccc2	11.42	abcd	205	abcd	
	dddd3	cccc1	89.45	abcd	30.0	abcd

Table Caption

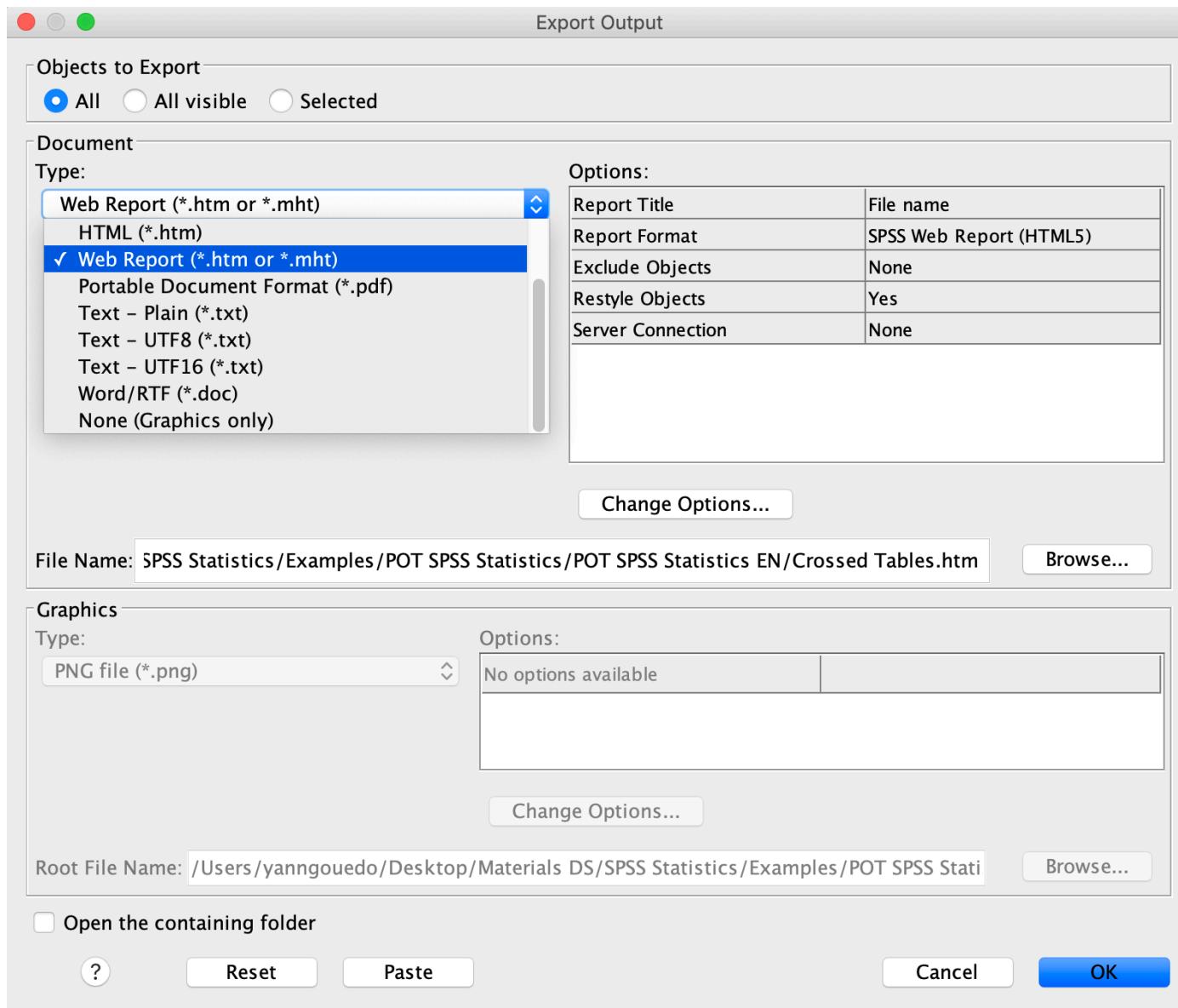
a.Text for footnote a.  
b.Text for footnote b.

Save Look Save As Edit Look

Socio_Professional_Cate	gory	Craftsmen, Shopkeepers and Businessmen (CSB)	Usage	Private & Professional	Alfa Romeo Count	Audi Count	Autres Count	BMW Count	Chrysler Count	Citroën Count	Fiat Count	Ford Count	H C
				Private & Work journey	2	0	0	0	5	17	3	3	3
			Employees (E)	Usage	Private	1	0	0	0	0	5	3	2
					Private & Professional	0	0	0	0	1	6	2	5
					Private & Work journey	0	0	0	0	0	3	1	3
			Farmers (F)	Usage	Private & Professional	4	3	2	4	3	39	14	16
			Intermediate professions (IP)	Usage	Private	2	2	0	1	3	25	2	6
					Private & Professional	3	2	2	1	2	28	8	8
					Private & Work journey	0	0	0	0	2	9	0	2
			Other people without professional activity (OPWPA)	Usage	Private	0	0	0	0	0	0	0	0
					Private & Professional	1	0	0	0	0	0	0	0
					Private & Work journey	0	0	1	1	0	2	0	6
			Retirees (R)	Usage	Private	5	0	0	2	9	33	10	9
			Senior Managers and Professional Occupations (SMPO)	Usage	Private	2	2	3	3	6	17	9	15
					Private & Professional	3	1	0	0	3	26	8	17
					Private & Work journey	5	4	1	4	13	103	37	54
			Workers (W)	Usage	Private & Professional	2	0	0	0	2	12	2	7

# IBM® SPSS® Statistics

## Web Report Exportation



# IBM® SPSS® Statistics

## Web Report Result

file:///Users/yanngouedo/Desktop/Materials DS/SPSS Statistics/Examples/POT SPSS Statistics/POT SPSS Statistics EN/Cro ...

Débuter avec Firefox TalkingData Fraud ... Air France IBM Forms Experien...

IBM SPSS Web Report - Output2

Not connected to the server

Contents Previous Next Help

Bar Chart

Bar Chart - Brand - October 11, 2019

Brand

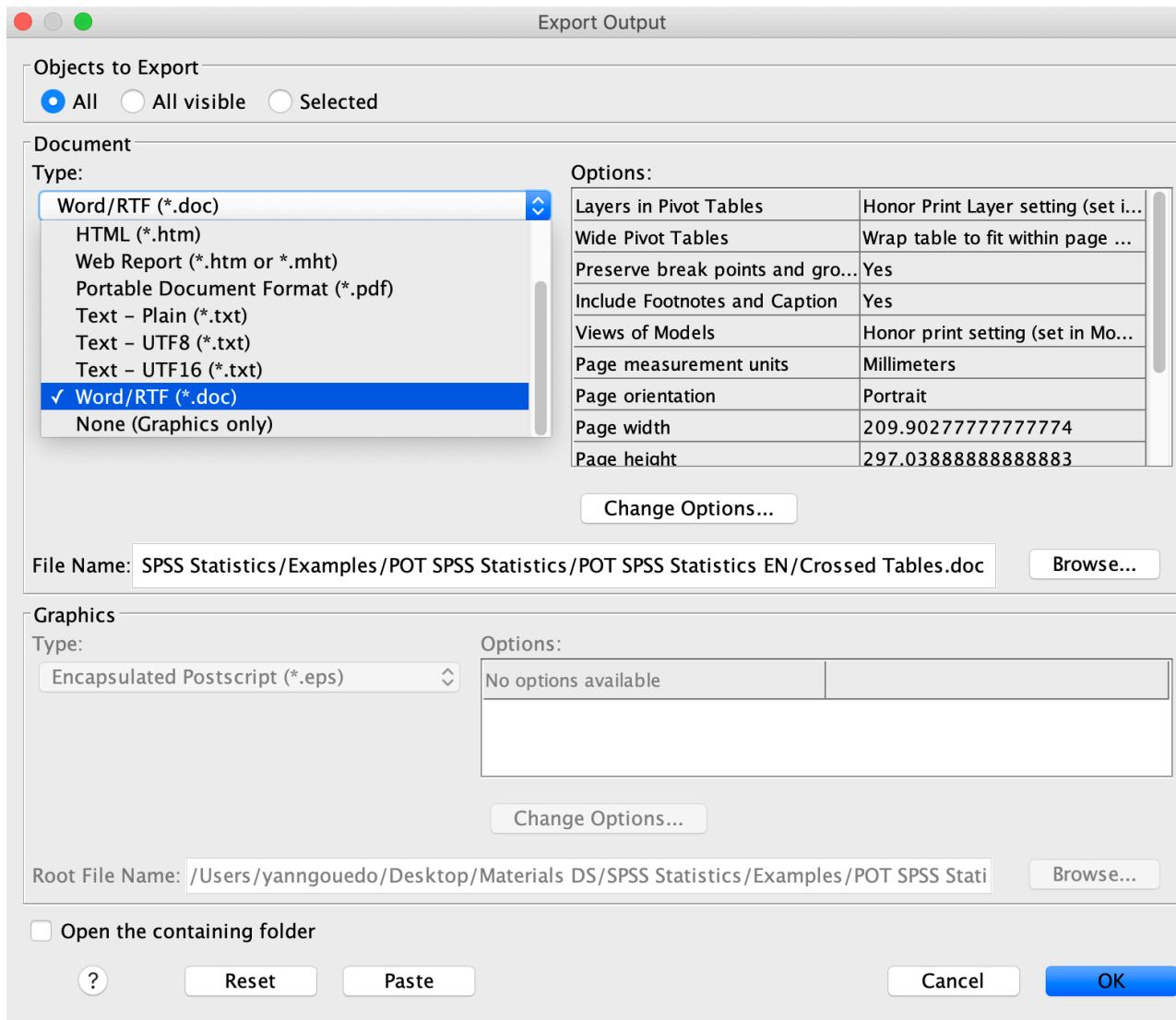
Frequency

Brand

Brand	Frequency
Alfa Romeo	~20
Audi	~10
Autres	~5
BMW	~5
Chrysler	~50
Citroën	~350
Fiat	~100
Ford	~150
Honda	~5
Kia	~5
Lancia	~5
Mazda	~5
Mercedes-Benz	~5
Nissan	~5
Porsche	~5
Peugeot	~400
Renault	~750
Saab	~50
Smart	~10
Toyota	~30
Volkswagen	~50
Volvo	~10

# IBM® SPSS® Statistics

## Other exportation choices



# IBM® SPSS® STATISTICS

## EXERCISE 5

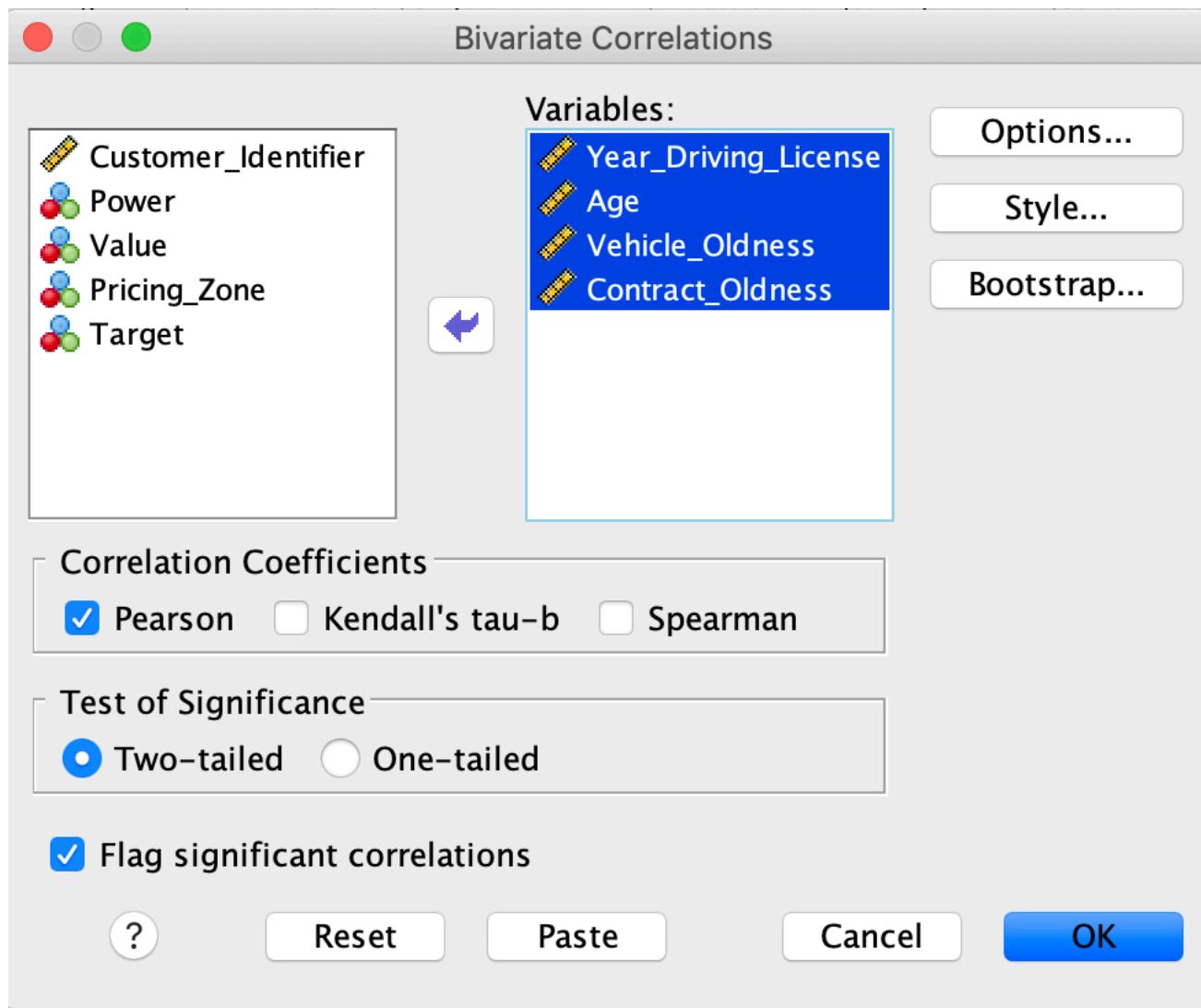
Correlation  
coefficients on  
variables

Mean comparison  
test to study the AGE  
distribution  
according to the 2  
populations SINISTER  
or absence of  
SINISTER

Clustering (two steps  
methodology)

# IBM® SPSS® Statistics

## Correlation coefficients on variables



# IBM® SPSS® Statistics Results

## Descriptives

Descriptive Statistics

	N	Mean	Std. Deviation
Year_Driving_License	2052	1986,65	12,115
Age	2007	48,18	14,561
Vehicle_Oldness	2160	10,75	5,112
Contract_Oldness	2160	8,92	5,759
Valid N (listwise)	2007		

## CORRELATIONS

```
/VARIABLES=Year_Driving_License Age Vehicle_Oldness Contract_Oldness
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

## Correlations

Correlations

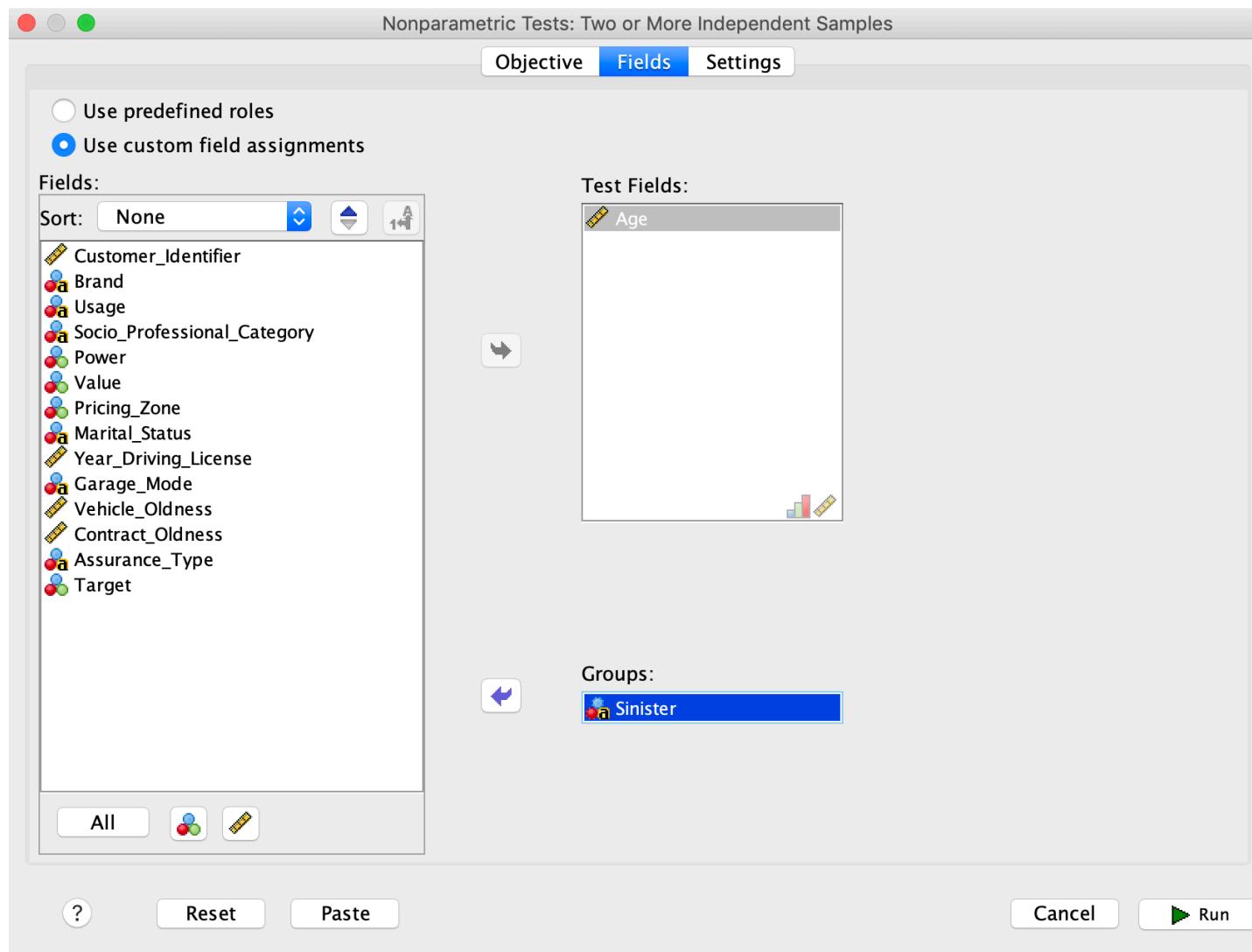
		Year_Driving_License	Age	Vehicle_Oldness	Contract_Oldness
Year_Driving_License	Pearson Correlation	1	-,881**	,011	-,366**
	Sig. (2-tailed)		,000	,627	,000
	N	2052	2007	2052	2052
Age	Pearson Correlation	-,881**	1	,024	,382**
	Sig. (2-tailed)	,000		,275	,000
	N	2007	2007	2007	2007
Vehicle_Oldness	Pearson Correlation	,011	,024	1	,044*
	Sig. (2-tailed)	,627	,275		,041
	N	2052	2007	2160	2160
Contract_Oldness	Pearson Correlation	-,366**	,382**	,044*	1
	Sig. (2-tailed)	,000	,000	,041	
	N	2052	2007	2160	2160

\*\*. Correlation is significant at the 0.01 level (2-tailed).

\*. Correlation is significant at the 0.05 level (2-tailed).

# IBM® SPSS® Statistics

## Mean comparison test to study the AGE distribution on the SINISTER (0,1) population



# IBM® SPSS® Statistics

## Results

### Nonparametric Tests

#### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Age is the same across categories of Sinister.	Independent-Samples Mann-Whitney U Test	,000	Reject the null hypothesis.

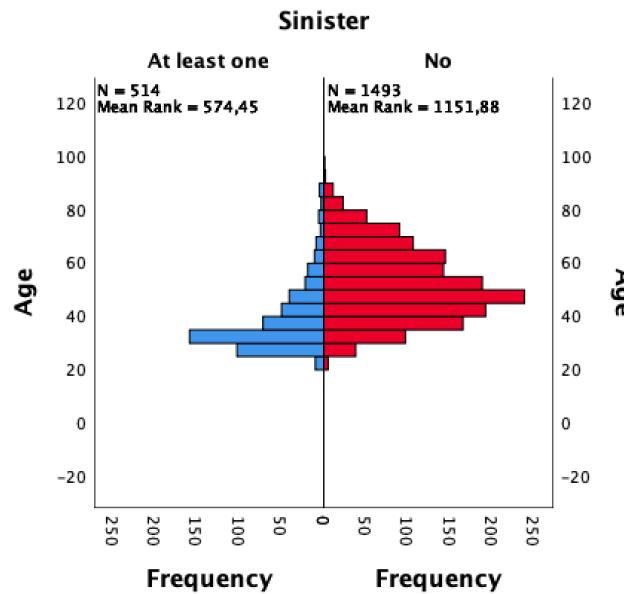
Asymptotic significances are displayed. The significance level is ,050.

#### Age across Sinister

##### Independent-Samples Mann-Whitney U Test Summary

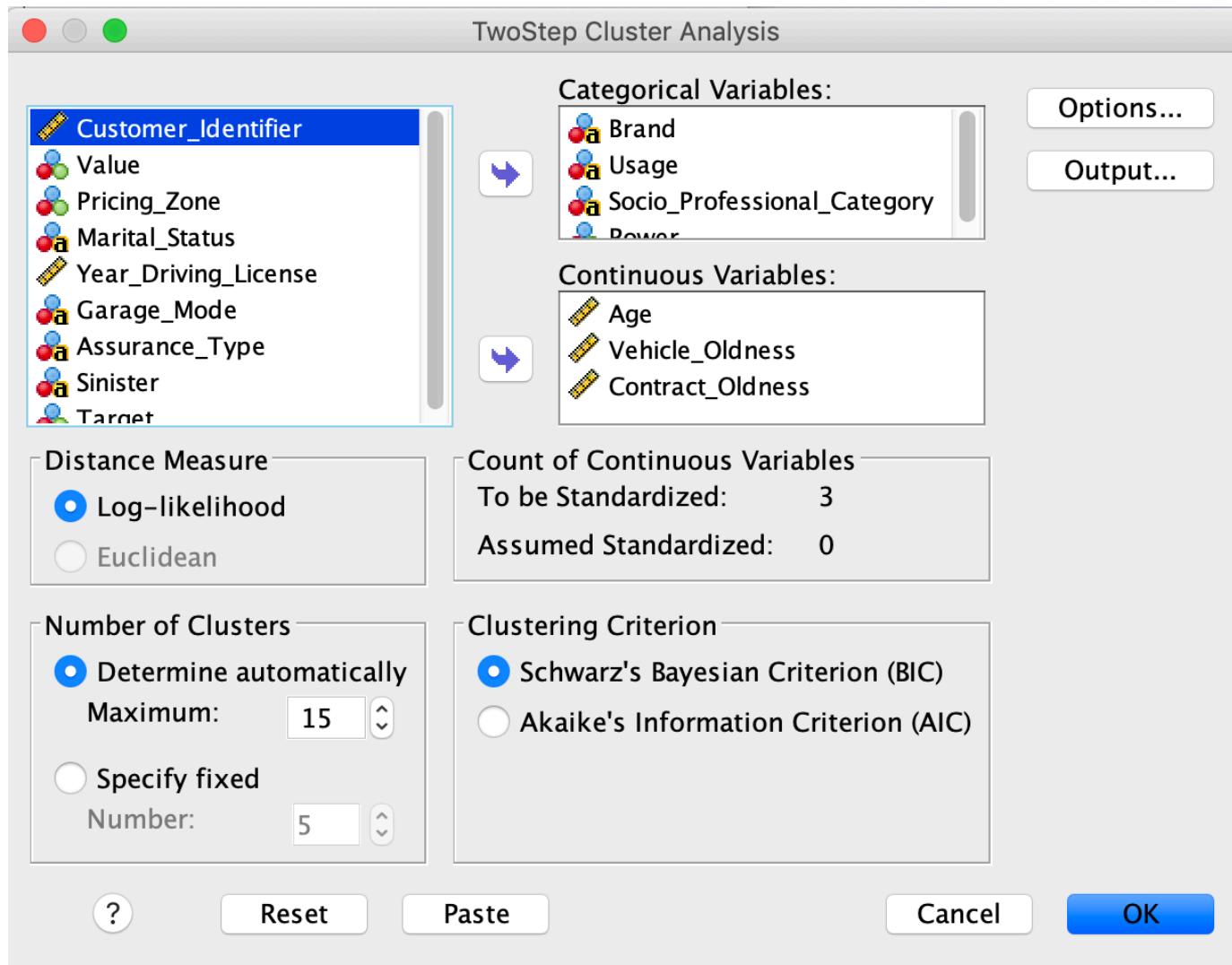
Total N	2007
Mann-Whitney U	604489,500
Wilcoxon W	1719760,500
Test Statistic	604489,500
Standard Error	11329,251
Standardized Test Statistic	19,488
Asymptotic Sig.(2-sided test)	,000

##### Independent-Samples Mann-Whitney U Test

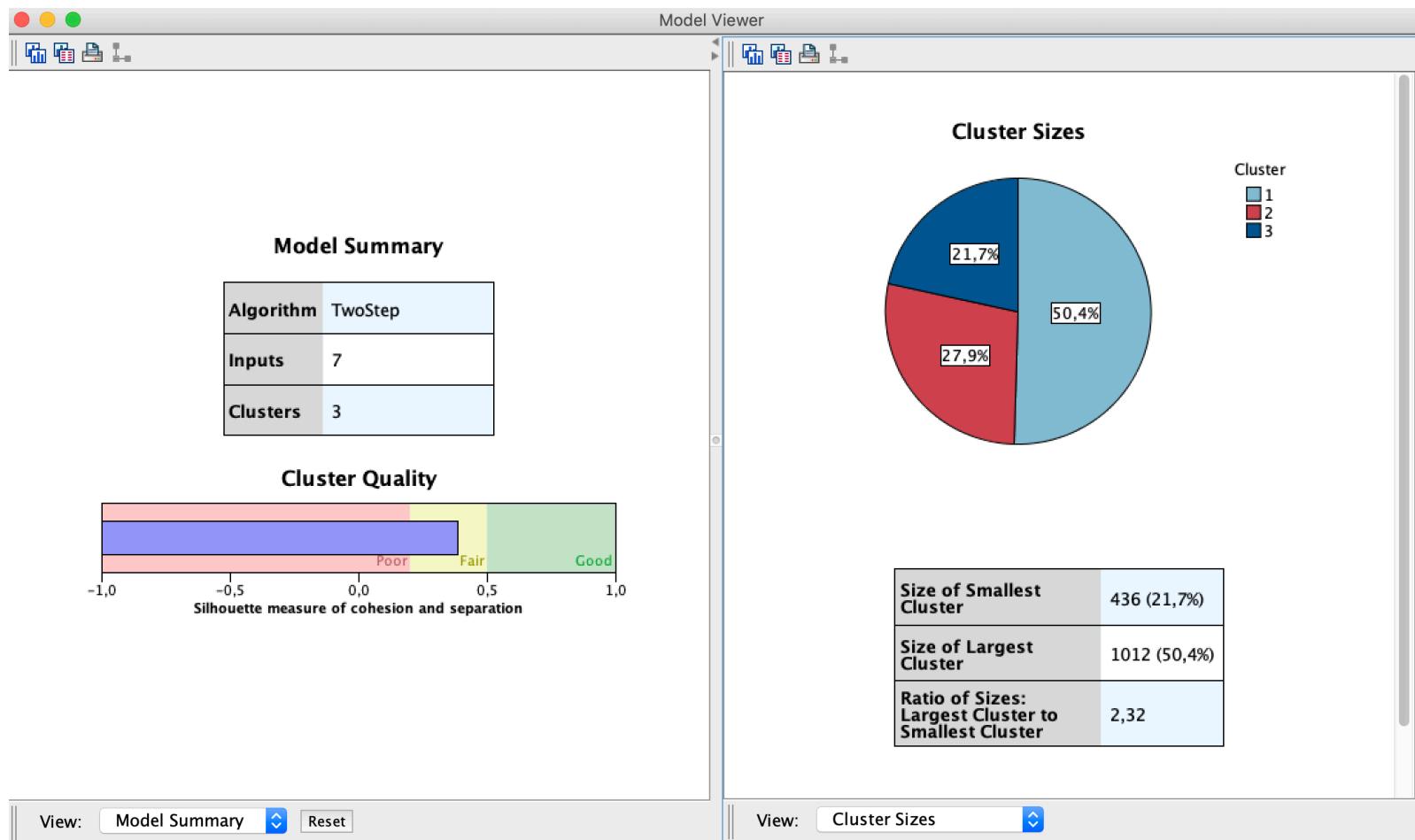


# IBM® SPSS® Statistics

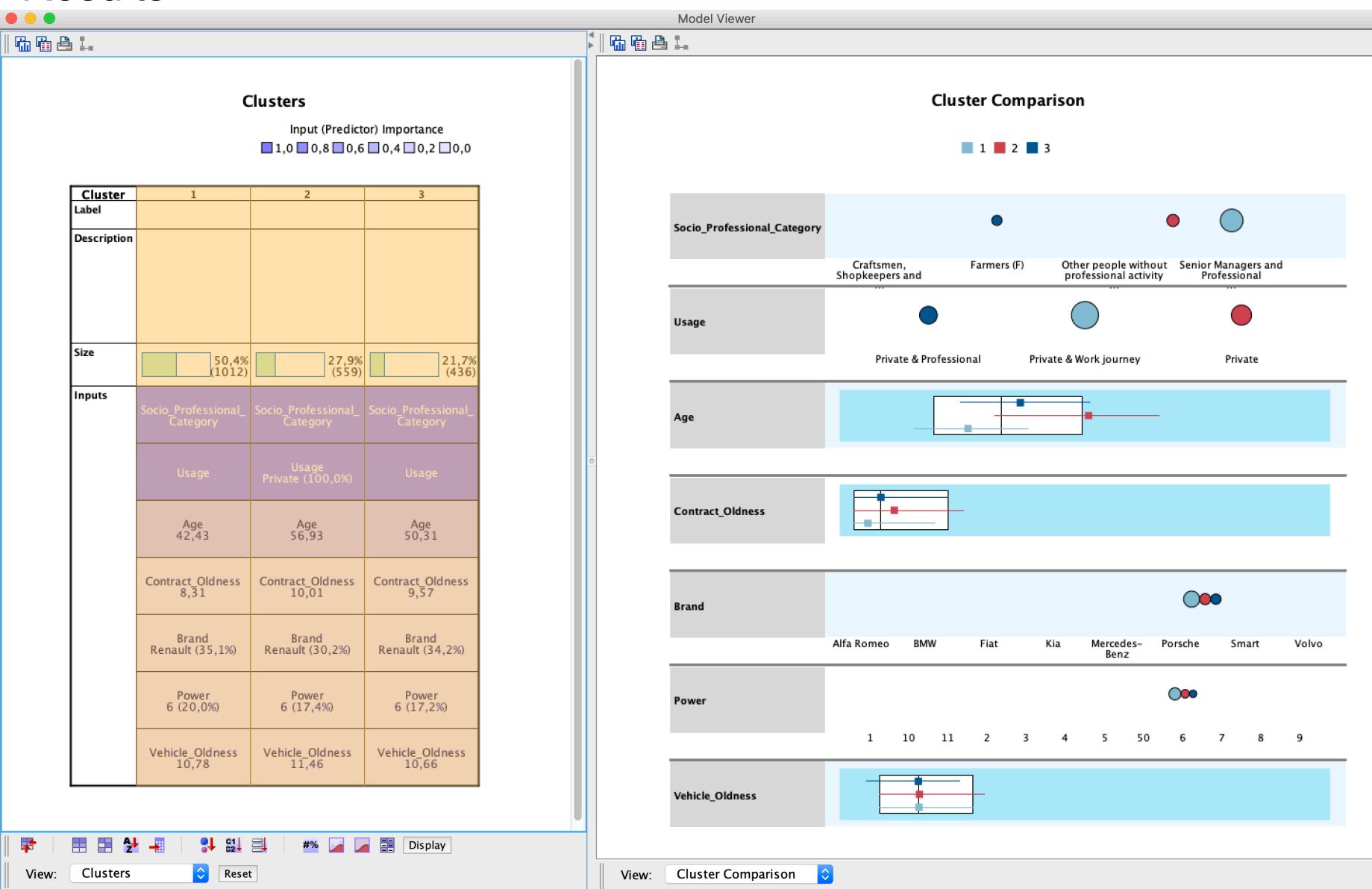
## Two-Steps clustering



# IBM® SPSS® Statistics Results



# IBM® SPSS® Statistics Results



# IBM® SPSS® STATISTICS

## EXERCISE 6

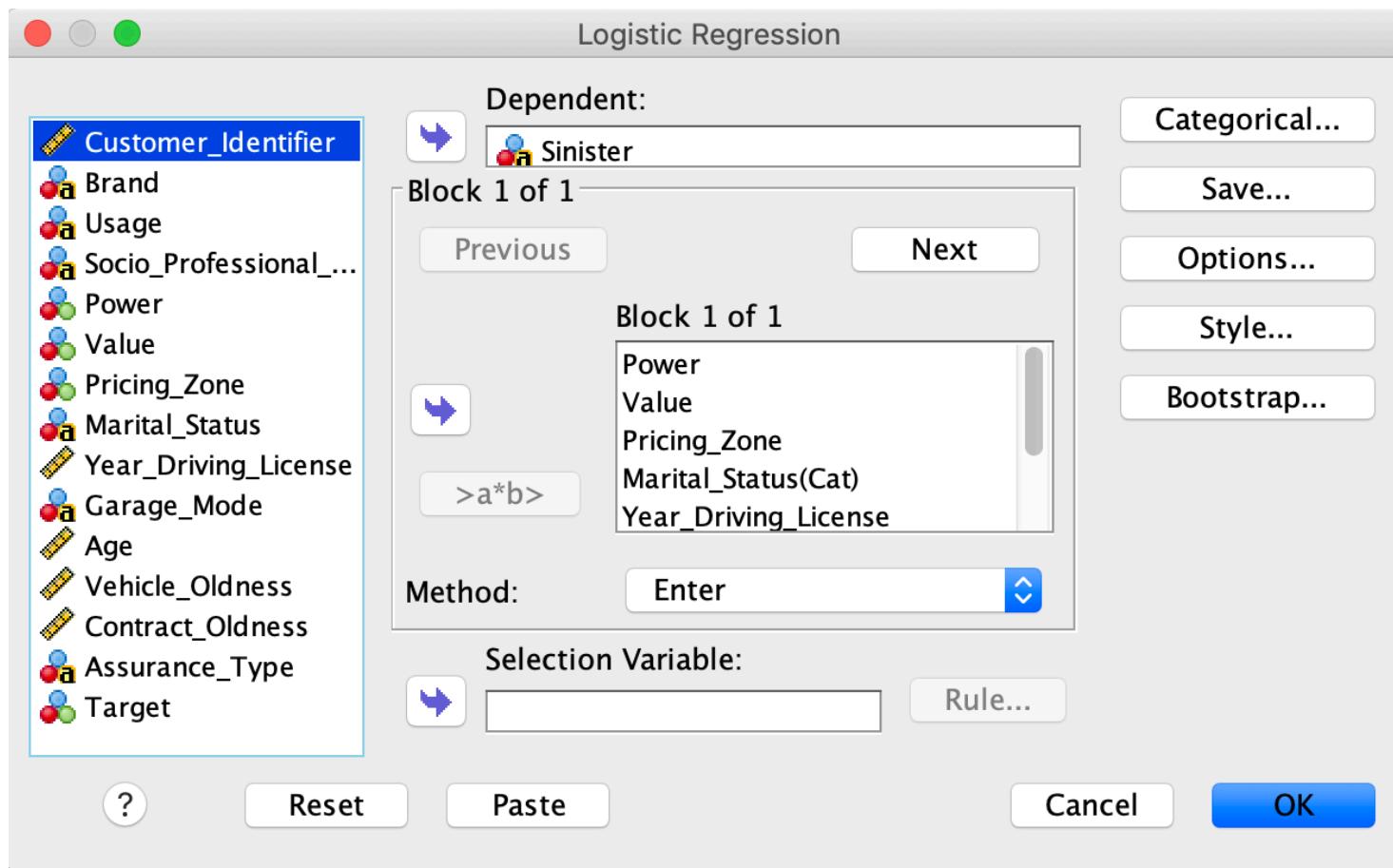
Predictive  
modelling on  
SINISTER

Logistic  
Regression

Decision Tree

# IBM® SPSS® Statistics

## Logistic Regression



# IBM® SPSS® Statistics Results

## Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	492,405	10	,000
	Block	492,405	10	,000
	Model	492,405	10	,000

## Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1779,365 <sup>a</sup>	,219	,322

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

## Classification Table<sup>a</sup>

Observed	Sinister	Predicted		Percentage Correct	
		Sinister			
		At least	No		
Step 1	Sinister	232	280	45,3	
	No	104	1378	93,0	
	Overall Percentage			80,7	

a. The cut value is ,500

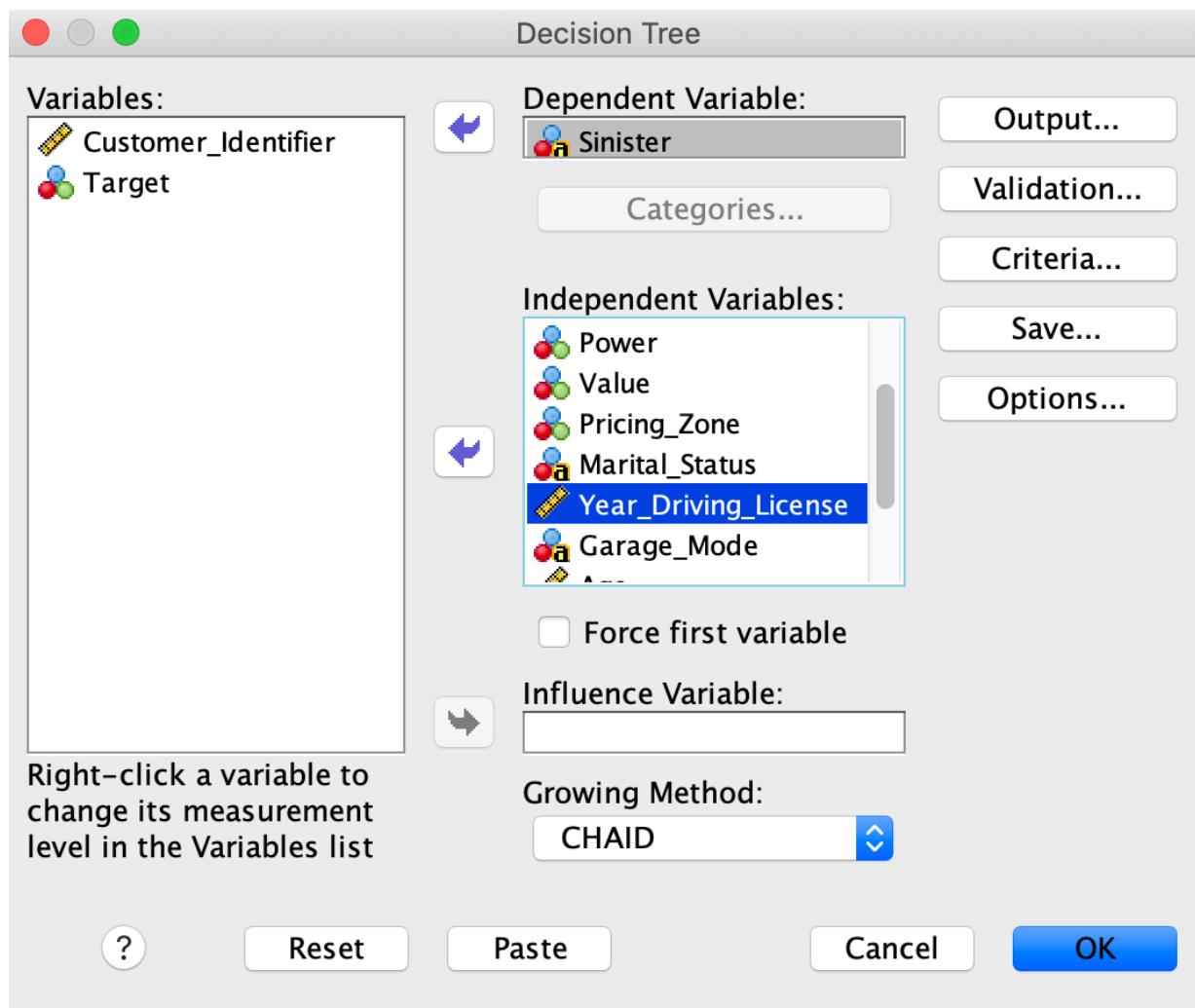
## Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	Power	,003	,036	,007	1	,935
	Value	-,010	,020	,267	1	,606
	Pricing_Zone	-,129	,030	18,060	1	,000
	Marital_Status			12,657	3	,005
	Marital_Status(1)	,127	,538	,056	1	,813
	Marital_Status(2)	,144	,223	,417	1	,518
	Marital_Status(3)	-,378	,241	2,464	1	,116
	Year_Driving_License	-,094	,012	62,485	1	,000
	Age	,007	,009	,630	1	,427
	Vehicle_Oldness	-,028	,012	5,066	1	,024
	Contract_Oldness	,035	,015	5,310	1	,021
	Constant	187,753	23,911	61,658	1	,000

a. Variable(s) entered on step 1: Power, Value, Pricing\_Zone, Marital\_Status, Year\_Driving\_License, Age, Vehicle\_Oldness, Contract\_Oldness.

# IBM® SPSS® Statistics

## Decision Tree

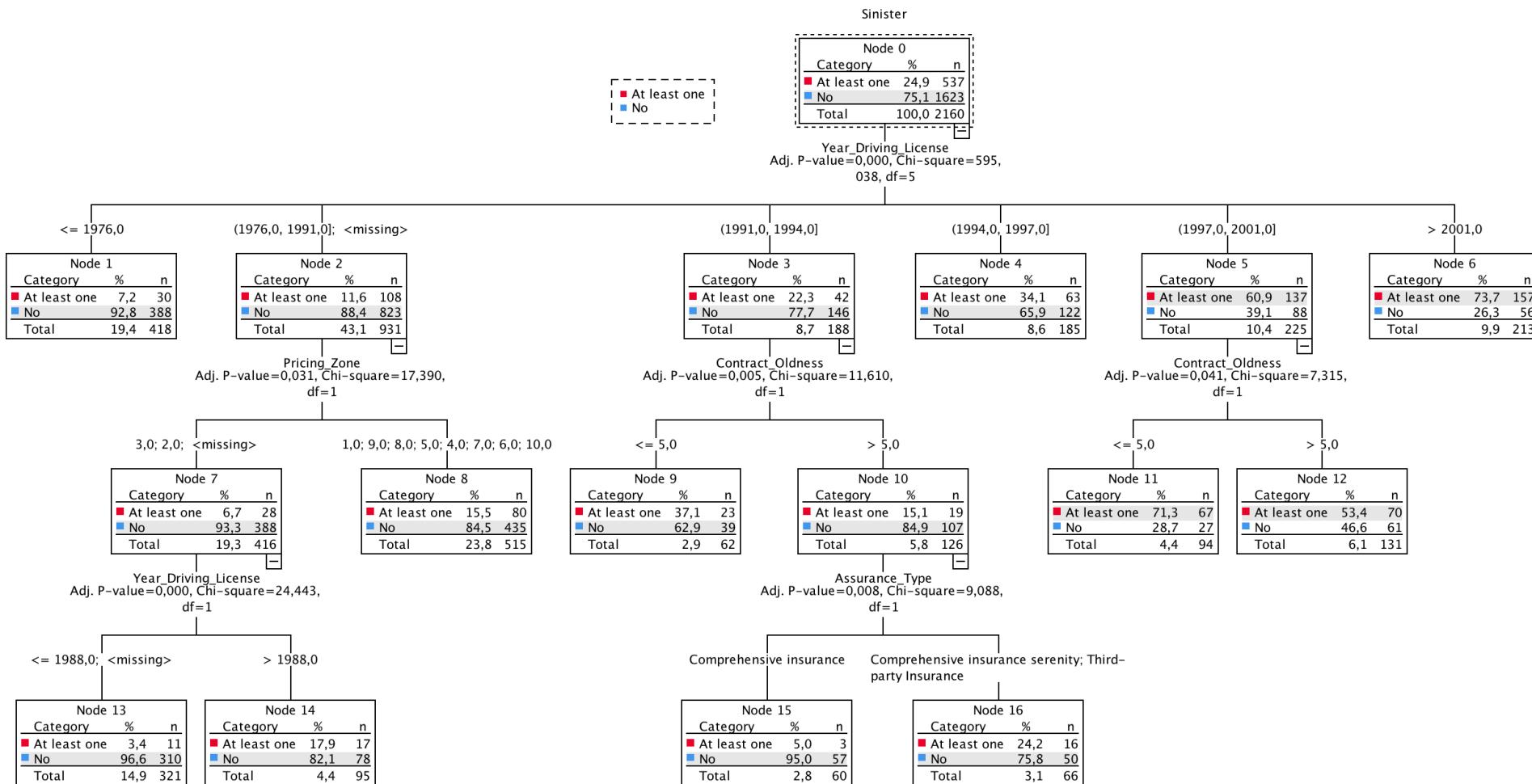


# IBM® SPSS® Statistics

## Results

Observed	Predicted		Percent Correct
	At least one	No	
At least one	294	243	54,7%
No	144	1479	91,1%
Overall Percentage	20,3%	79,7%	82,1%

Growing Method: CHAID



# IBM® SPSS® STATISTICS

## EXERCISE 7

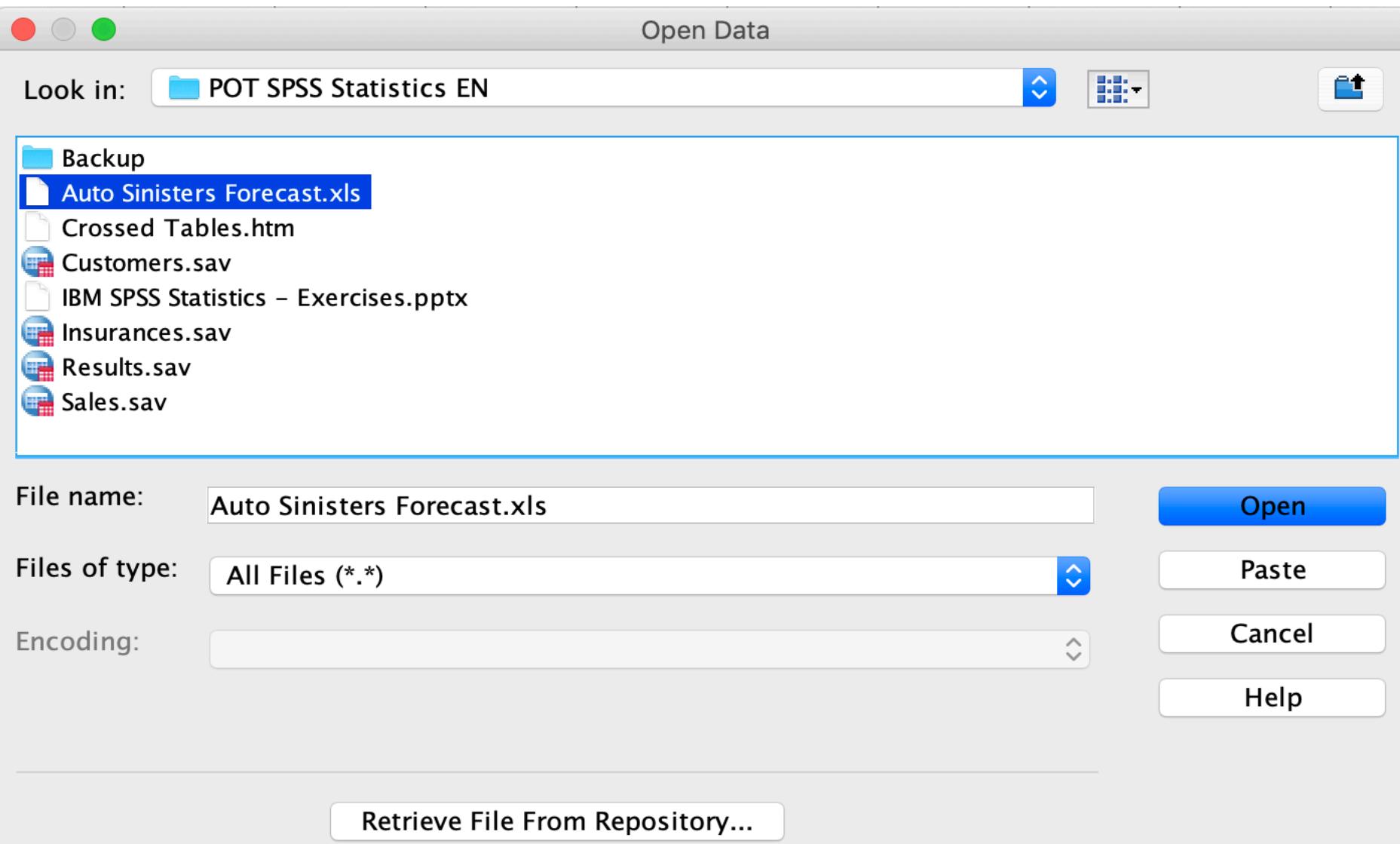
Auto Sinisters  
Forecast.xls file  
upload

Time series  
Visualization

Time series  
modelling

# IBM® SPSS® Statistics

## Auto Sinisters Forecast.xls file upload



# IBM® SPSS® Statistics Data

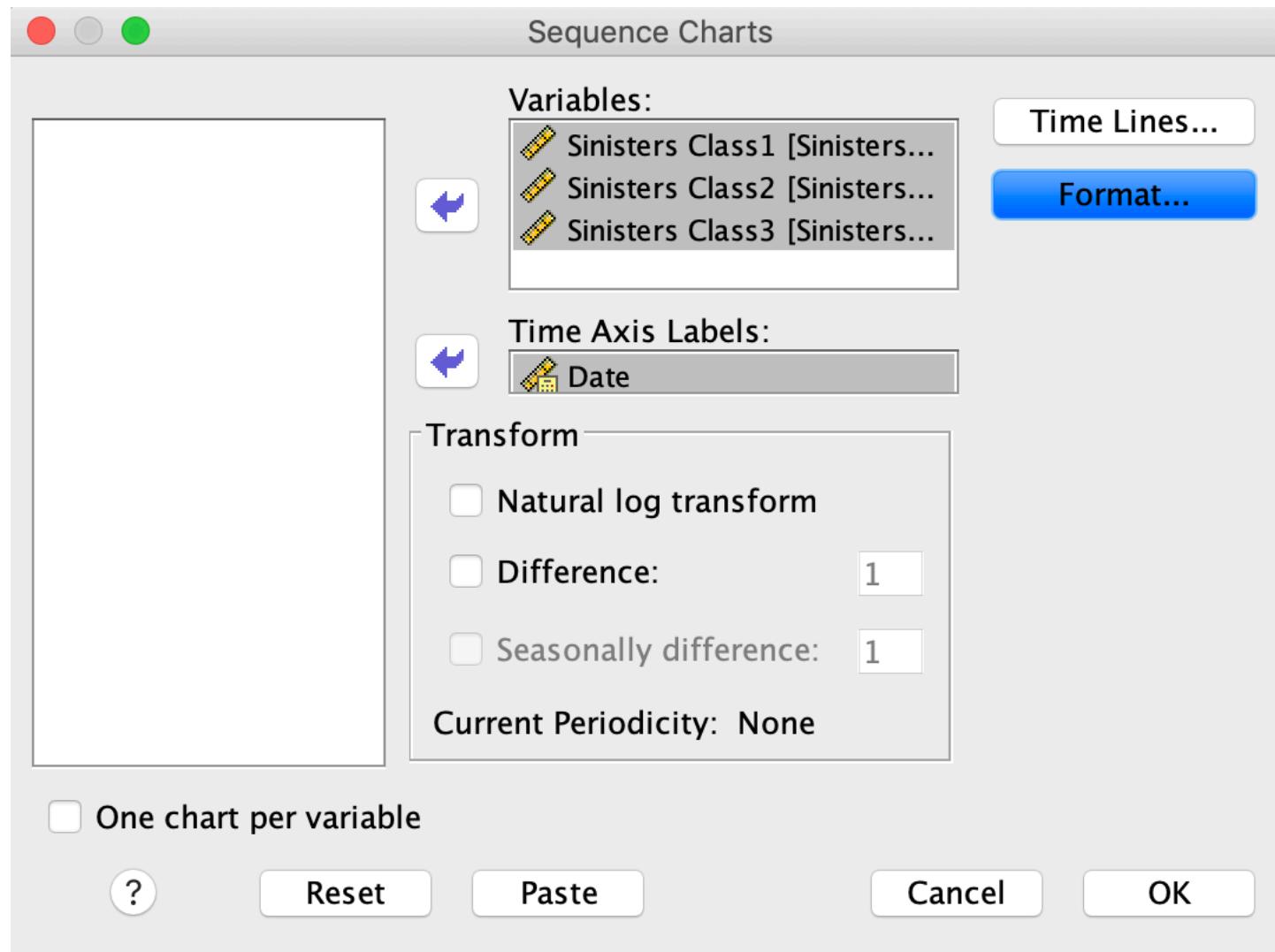
Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

Visible: 4 of 4 Variables

	Date	SinistersClas s1	SinistersClas s2	SinistersClas s3	var							
1	01-Aug-2010	16579	10776	5400								
2	01-Sep-2010	18236	10822	7500								
3	01-Oct-2010	43394	22846	7500								
4	01-Nov-2010	30908	11103	9600								
5	01-Dec-2010	28702	16067	15900								
6	01-Jan-2011	29648	11061	11100								
7	01-Feb-2011	31142	11329	11100								
8	01-Mar-2011	31177	16789	10800								
9	01-Apr-2011	30672	14453	7500								
10	01-May-2011	37633	16056	5400								
11	01-Jun-2011	33891	24557	8100								
12	01-Jul-2011	51378	34646	12000								
13	01-Aug-2011	18103	11528	14700								
14	01-Sep-2011	20980	12885	17700								
15	01-Oct-2011	34503	14748	16500								
16	01-Nov-2011	26784	9595	19800								
17	01-Dec-2011	31790	15926	21900								
18	01-Jan-2012	32433	11383	16800								
19	01-Feb-2012	37180	16053	9900								
20	01-Mar-2012	29659	14803	11400								
21	01-Apr-2012	33238	14125	13800								

# IBM® SPSS® Statistics

## Times Series



# IBM® SPSS® Statistics Results

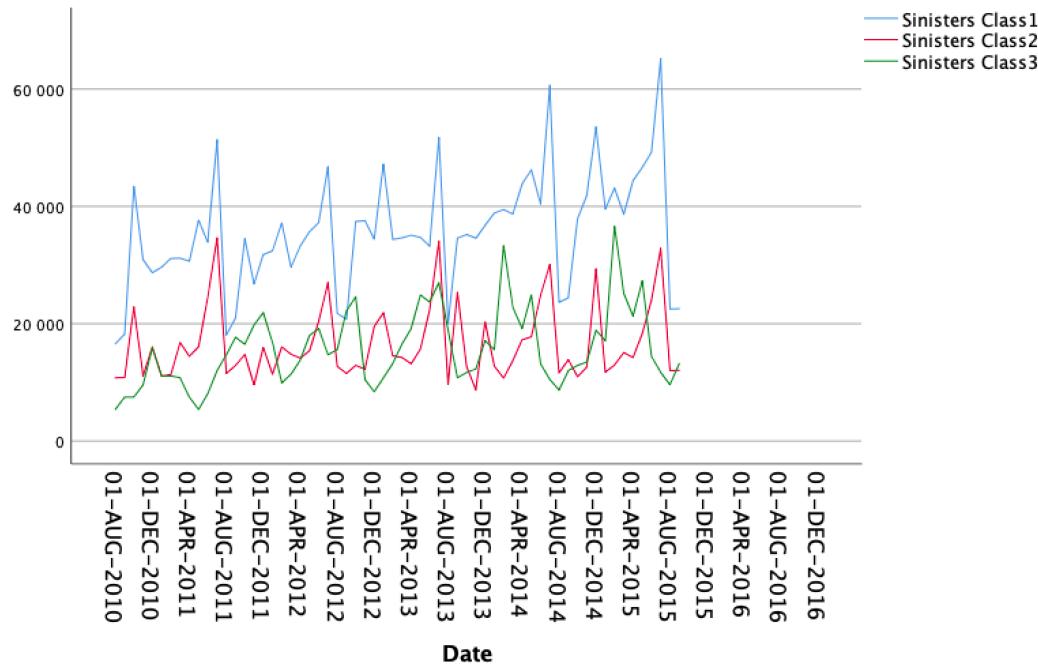
## Case Processing Summary

	Sinisters Class1	Sinisters Class2	Sinisters Class3
Series or Sequence Length	77	77	77
Number of Missing Values in the Plot	User-Missing	0	0
	System-Missing	15	15

## Model Description

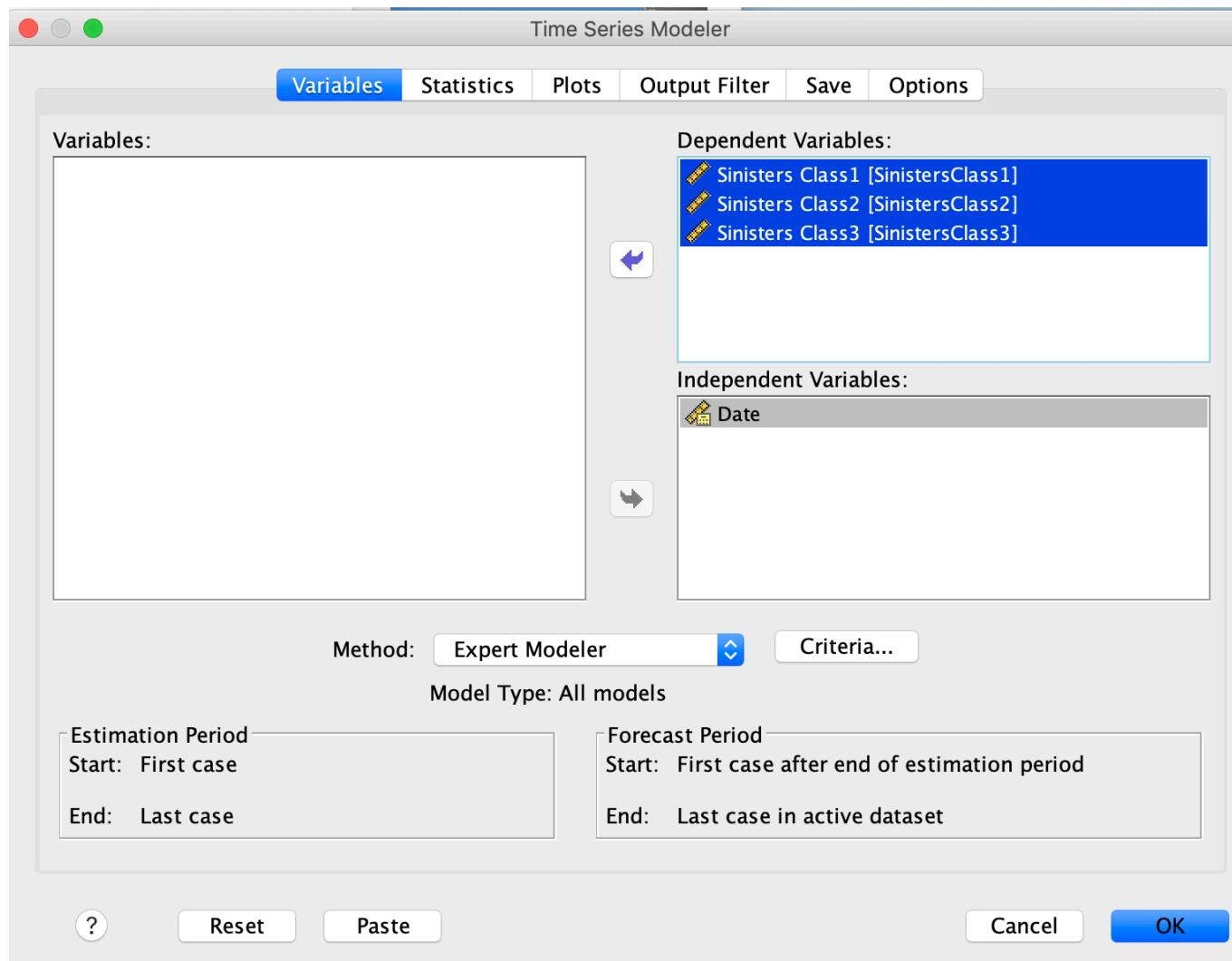
Model Name	MOD_1
Series or Sequence	1
	Sinisters Class1
	2
	Sinisters Class2
	3
	Sinisters Class3
Transformation	None
Non-Seasonal Differencing	0
Seasonal Differencing	0
Length of Seasonal Period	No periodicity
Horizontal Axis Labels	Date
Intervention Onsets	None
For Each Observation	Values not joined

Applying the model specifications from MOD\_1

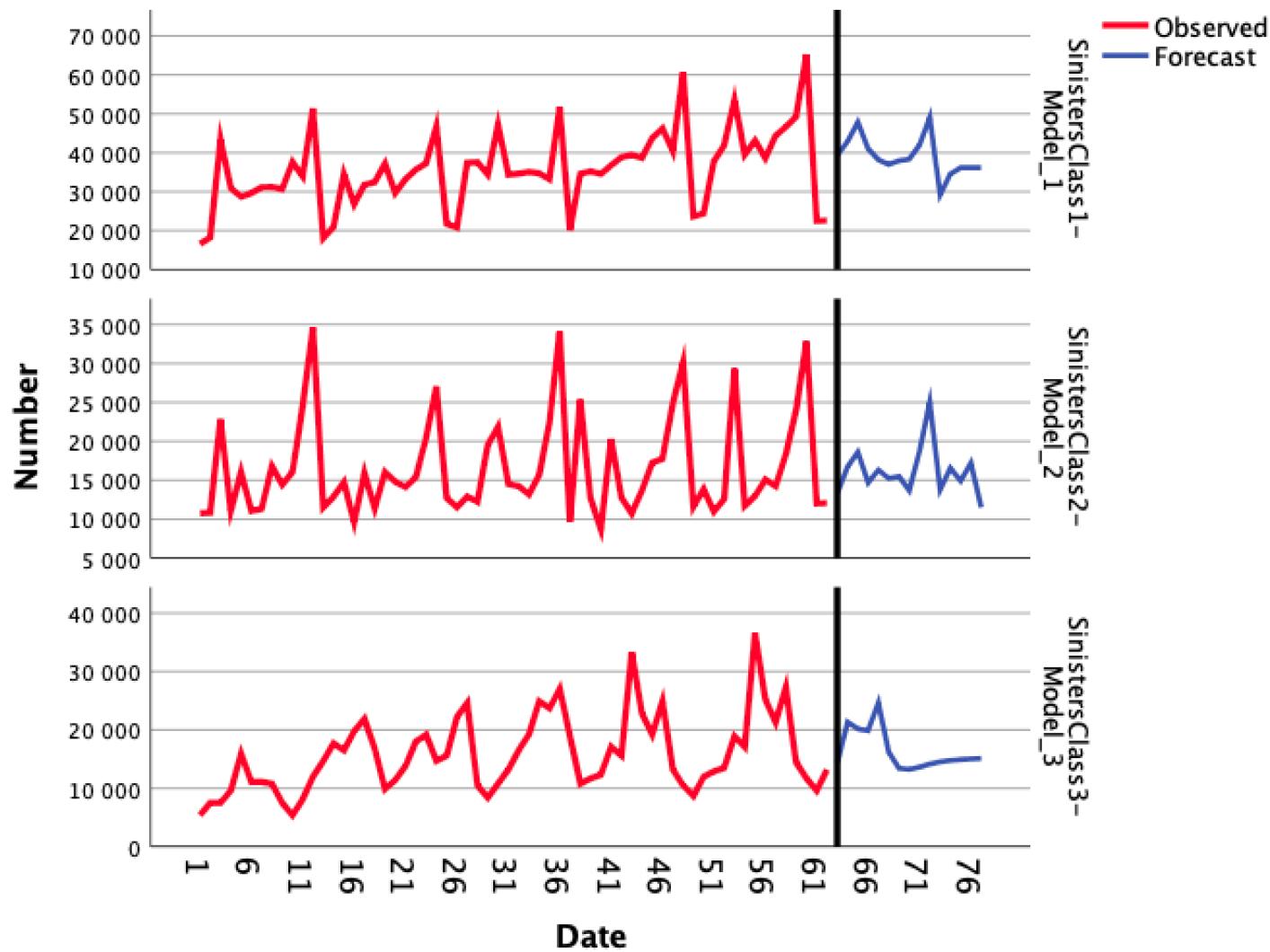


# IBM® SPSS® Statistics

## Time series modelling



# IBM® SPSS® Statistics Results



# IBM® SPSS® STATISTICS

## EXERCISE 8

RFM  
Analysis

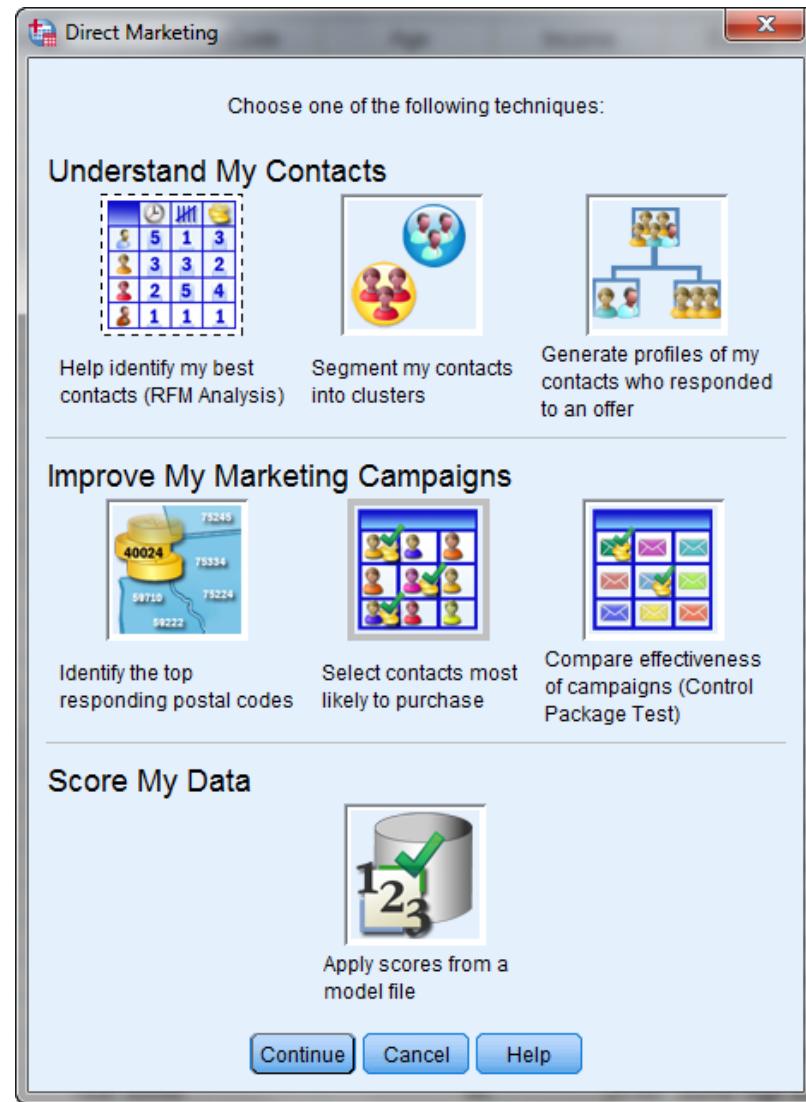
Generate  
profiles

Top postal  
codes

# IBM® SPSS® Statistics

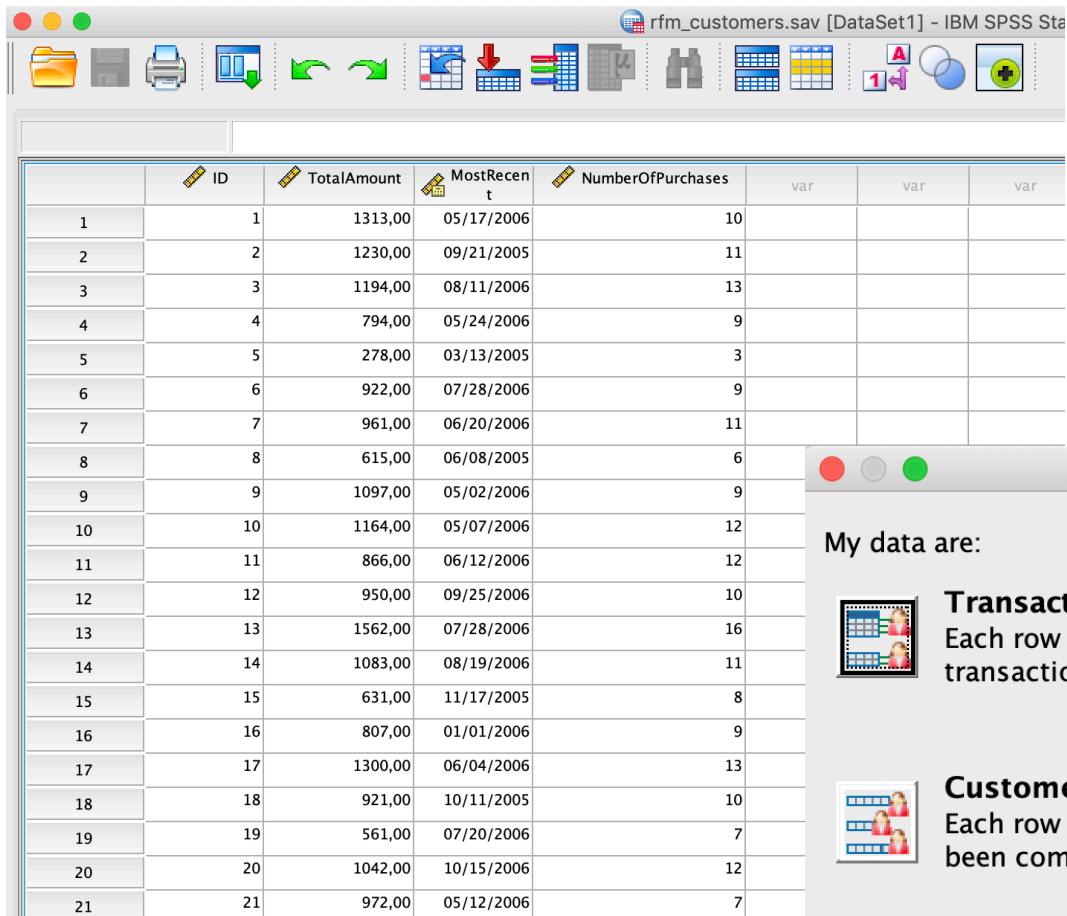
## Direct Marketing Analysis

- RFM analysis (rfm\_customers.sav or rfm\_transactions.sav)
- Segment customers (two-step clustering) (dmdata\_historical.sav)
- Generate profiles (dmdata\_historical.sav)
- Top postal codes (dmdata\_historical.sav)
- Select contacts (logistic regression and scoring) (dmdata\_historical.sav)



# IBM® SPSS® Statistics

## RFM analysis (customer)



	ID	TotalAmount	MostRecent	NumberOfPurchases	var	var	var
1	1	1313,00	05/17/2006		10		
2	2	1230,00	09/21/2005		11		
3	3	1194,00	08/11/2006		13		
4	4	794,00	05/24/2006		9		
5	5	278,00	03/13/2005		3		
6	6	922,00	07/28/2006		9		
7	7	961,00	06/20/2006		11		
8	8	615,00	06/08/2005		6		
9	9	1097,00	05/02/2006		9		
10	10	1164,00	05/07/2006		12		
11	11	866,00	06/12/2006		12		
12	12	950,00	09/25/2006		10		
13	13	1562,00	07/28/2006		16		
14	14	1083,00	08/19/2006		11		
15	15	631,00	11/17/2005		8		
16	16	807,00	01/01/2006		9		
17	17	1300,00	06/04/2006		13		
18	18	921,00	10/11/2005		10		
19	19	561,00	07/20/2006		7		
20	20	1042,00	10/15/2006		12		
21	21	972,00	05/12/2006		7		
22	22	845,00	09/29/2005		10		

RFM Analysis: Data Format

My data are:



### Transaction data

Each row contains data for one transaction. For the analysis, transactions will be combined by customer identifiers.



### Customer data

Each row contains data for one customer. The data have already been combined by customer over transactions.

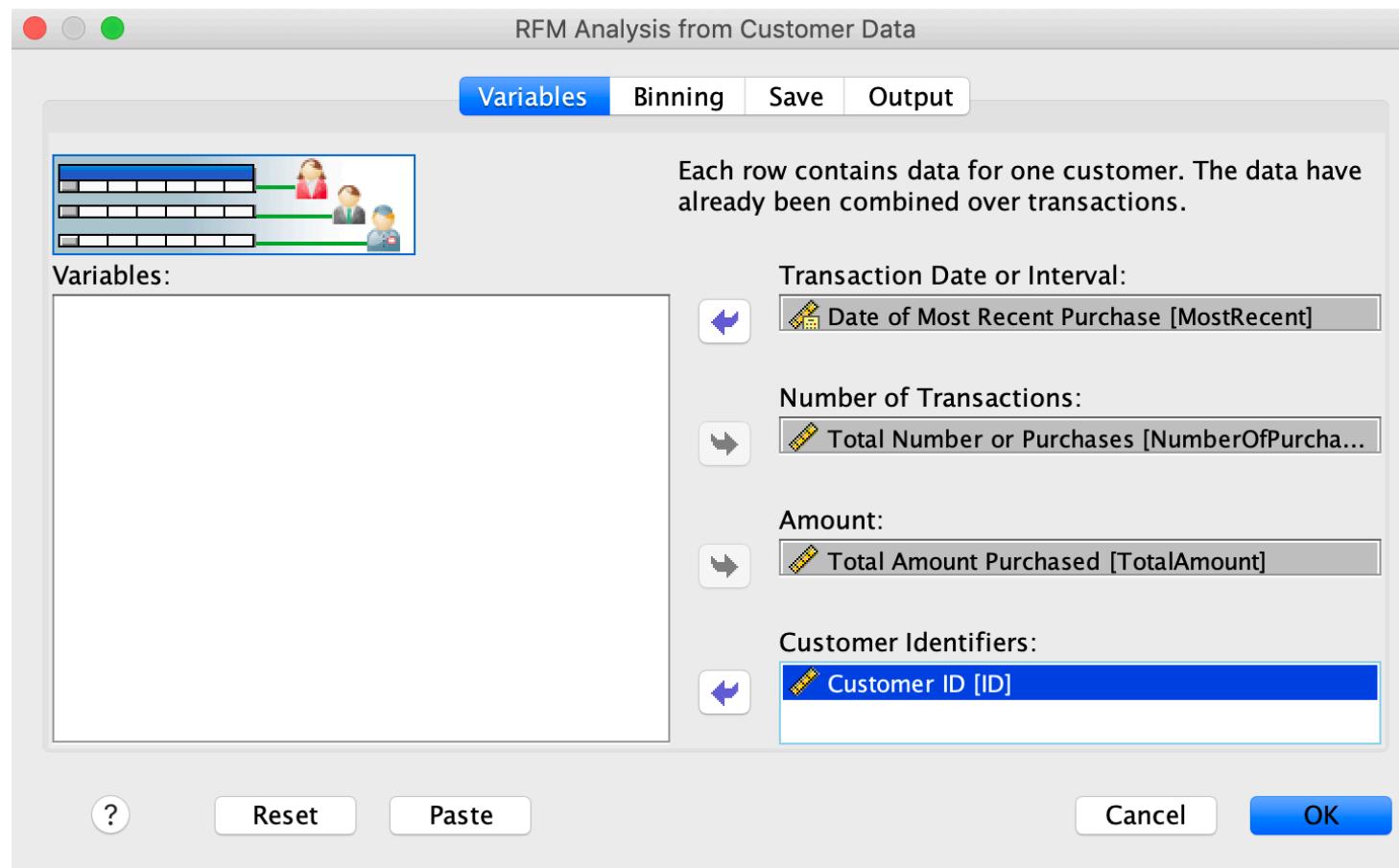


Cancel

Continue

# IBM® SPSS® Statistics

## RFM analysis (customer)



# IBM® SPSS® Statistics Results

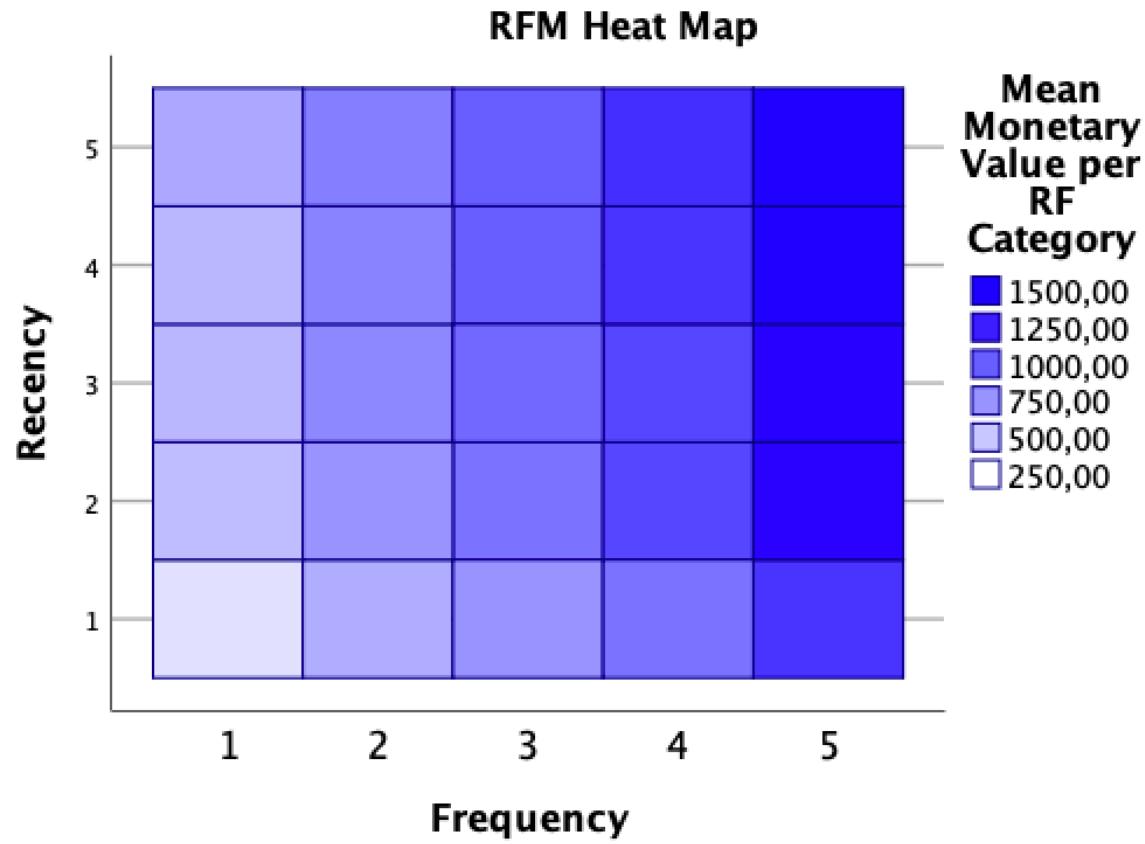
rfm\_customers.sav [DataSet1] - IBM SPSS Statistics Data Editor

Visible: 8 of 8 Variables

	ID	TotalAmount	MostRecent	NumberOfPurchases	Recency_score	Frequency_score	Monetary_score	RFM_score	var	var	var
1	1	1313,00	05/17/2006	10	2	3	5	235			
2	2	1230,00	09/21/2005	11	1	5	4	154			
3	3	1194,00	08/11/2006	13	3	5	2	352			
4	4	794,00	05/24/2006	9	2	3	2	232			
5	5	278,00	03/13/2005	3	1	1	2	112			
6	6	922,00	07/28/2006	9	3	2	4	324			
7	7	961,00	06/20/2006	11	2	4	2	242			
8	8	615,00	06/08/2005	6	1	2	3	123			
9	9	1097,00	05/02/2006	9	2	3	5	235			
10	10	1164,00	05/07/2006	12	2	4	4	244			
11	11	866,00	06/12/2006	12	2	4	1	241			
12	12	950,00	09/25/2006	10	4	3	2	432			
13	13	1562,00	07/28/2006	16	3	5	4	354			
14	14	1083,00	08/19/2006	11	3	4	3	343			
15	15	631,00	11/17/2005	8	1	3	1	131			
16	16	807,00	01/01/2006	9	1	4	2	142			
17	17	1300,00	06/04/2006	13	2	5	3	253			
18	18	921,00	10/11/2005	10	1	4	3	143			
19	19	561,00	07/20/2006	7	3	1	3	313			
20	20	1042,00	10/15/2006	12	4	4	2	442			
21	21	972,00	05/12/2006	7	2	1	5	215			
22	22	845,00	09/29/2005	10	1	4	2	142			
23	23	933,00	03/28/2006	11	1	5	1	151			
24	24	1189,00	10/13/2006	16	4	5	1	451			
25	25	669,00	05/28/2006	10	2	3	1	231			
26	26	344,00	08/02/2005	6	1	2	1	121			
27	27	581,00	05/21/2006	7	2	1	3	213			
28	28	530,00	08/31/2006	7	4	1	2	412			
29	29	544,00	02/21/2006	6	1	2	2	122			
30	30	1337,00	09/22/2006	13	4	4	4	444			
31	31	722,00	08/13/2006	8	3	2	2	322			
32	32	1046,00	07/10/2006	12	3	4	3	343			
--	--	1022,00	08/13/2006	11	2	4	2	242			

# IBM® SPSS® Statistics Results

## RFM Analysis



# IBM® SPSS® Statistics

## Generate profiles

dmdata\_historical.sav [DataSet5] - IBM SPSS Statistics Data Editor

Visible: 10 of 10 Variables

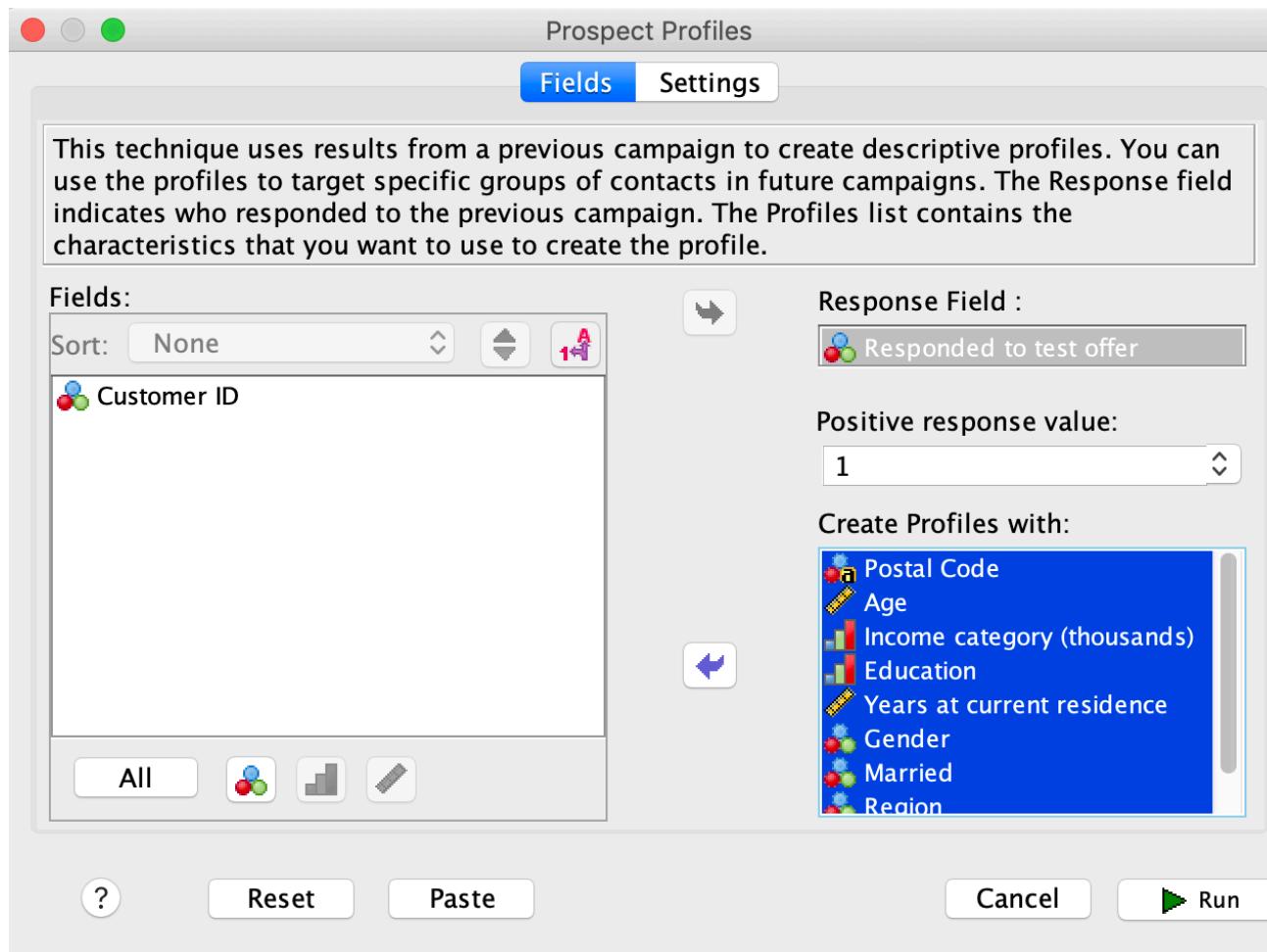
	ID	Responded	PostalCode	Age	Income	Education	Reside	Gender	Married	Region	var	var	var	var
3	01480	0	92590	56	3	2	11	1	0	4				
4	06118	0	92670	55	2	3	8	1	0	4				
5	07378	0	92690	56	2	1	6	1	0	4				
6	08467	0	93410	50	1	1	8	1	0	4				
7	09621	0	93480	67	4	2	9	1	0	4				
8	03029	0	93490	47	1	2	10	1	0	4				
9	02660	0	93490	58	4	5	11	1	0	4				
10	03179	0	93640	38	3	4	11	1	0	4				
11	03647	1	93760	27	2	5	14	1	0	4				
12	01741	0	93850	52	1	5	12	1	0	4				
13	05388	0	93900	66	3	4	10	1	0	4				
14	01942	0	93900	41	4	5	11	1	0	4				
15	06254	0	94120	48	4	4	12	1	0	4				
16	02164	0	94130	46	1	5	9	1	0	4				
17	02865	1	94150	37	4	3	10	1	0	4				
18	03330	0	94270	43	3	2	13	1	0	4				
19	06177	0	94330	59	4	1	10	1	0	4				
20	01554	0	94370	46	2	1	9	1	0	4				
21	06304	0	94510	53	2	1	12	1	0	4				
22	00673	0	94600	40	4	5	6	1	0	4				
23	05590	0	94630	48	1	3	10	1	0	4				
24	01504	0	94630	32	1	3	9	1	0	4				
25	06952	0	94670	56	3	2	10	1	0	4				
26	08089	0	94690	54	4	4	9	1	0	4				
27	03175	0	94800	55	2	1	6	1	0	4				
28	08891	0	94800	41	4	3	12	1	0	4				
29	05468	1	94810	39	4	2	10	1	0	4				
30	01767	0	94890	68	2	4	13	1	0	4				
31	09131	0	95000	53	1	2	8	1	0	4				
32	01913	0	95030	61	4	4	11	1	0	4				
33	05135	0	95080	61	1	3	7	1	0	4				
34	07533	0	95200	56	3	5	7	1	0	4				
35	02696	0	95270	55	4	5	9	1	0	4				

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

# IBM® SPSS® Statistics

## Generate profiles



# IBM® SPSS® Statistics Results

## Target Category: Yes

### Response Rate

Number	Description	Profile		Cumulative Response Rate
		Group Size	Response Rate	
1	Region = "West", "South", "East" Gender = "Female" Married = "No"	379	9,23%	9,23%
2	Region = "West", "South", "East" Gender = "Female" Married = "Yes"	299	5,02%	7,37%
3	Region = "West", "South", "East" Gender = "Male"	722	4,71%	6,00%
4	Region = "North"	517	2,51%	5,06%

# IBM® SPSS® Statistics

## Top postal codes

dmadata\_historical.sav [DataSet5] - IBM SPSS Statistics Data Editor

Visible: 10 of 10 Variables

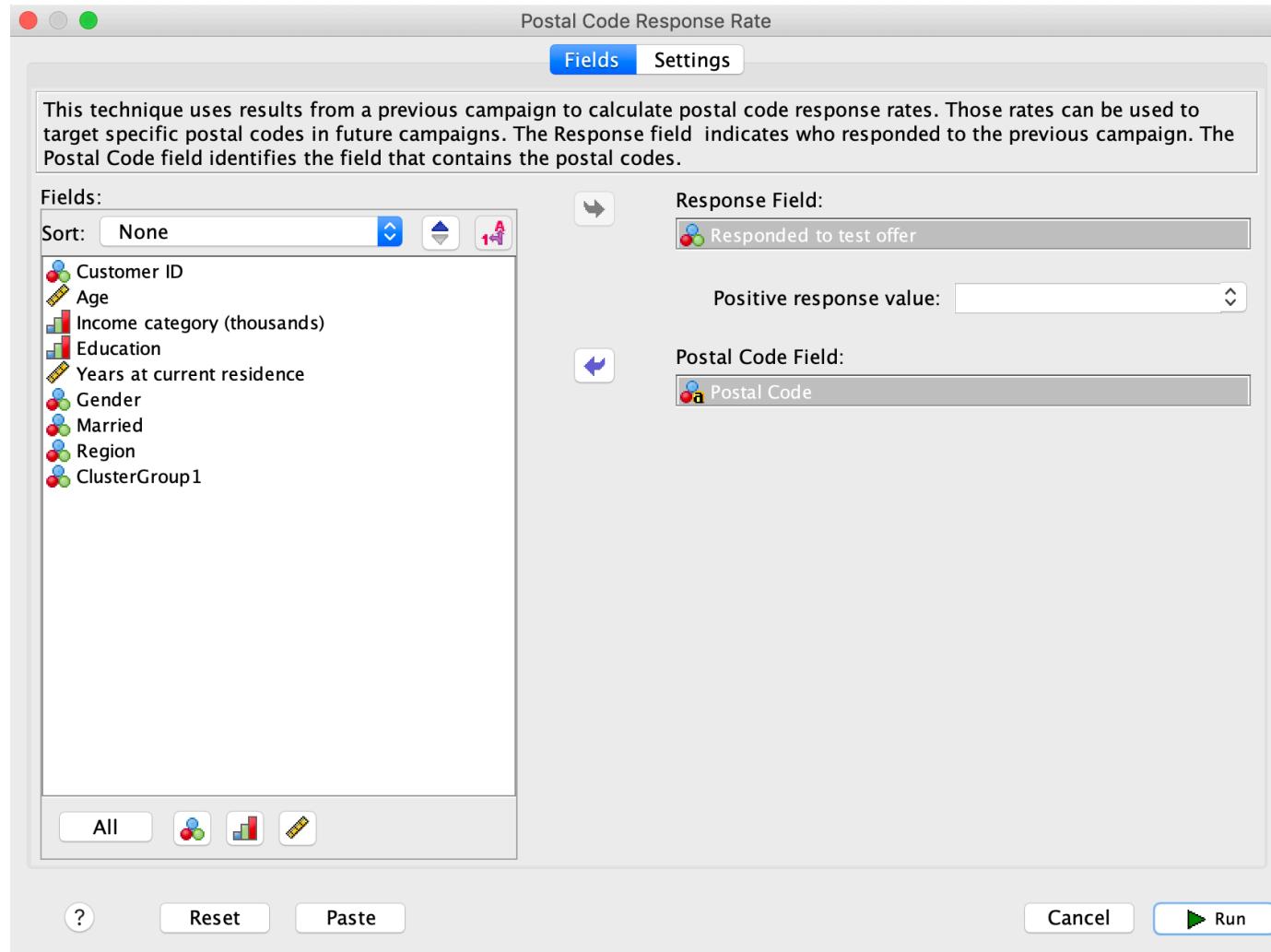
	ID	Responded	PostalCode	Age	Income	Education	Reside	Gender	Married	Region	var	var	var	var
3	01480	0	92590	56	3	2	11	1	0	4				
4	06118	0	92670	55	2	3	8	1	0	4				
5	07378	0	92690	56	2	1	6	1	0	4				
6	08467	0	93410	50	1	1	8	1	0	4				
7	09621	0	93480	67	4	2	9	1	0	4				
8	03029	0	93490	47	1	2	10	1	0	4				
9	02660	0	93490	58	4	5	11	1	0	4				
10	03179	0	93640	38	3	4	11	1	0	4				
11	03647	1	93760	27	2	5	14	1	0	4				
12	01741	0	93850	52	1	5	12	1	0	4				
13	05388	0	93900	66	3	4	10	1	0	4				
14	01942	0	93900	41	4	5	11	1	0	4				
15	06254	0	94120	48	4	4	12	1	0	4				
16	02164	0	94130	46	1	5	9	1	0	4				
17	02865	1	94150	37	4	3	10	1	0	4				
18	03330	0	94270	43	3	2	13	1	0	4				
19	06177	0	94330	59	4	1	10	1	0	4				
20	01554	0	94370	46	2	1	9	1	0	4				
21	06304	0	94510	53	2	1	12	1	0	4				
22	00673	0	94600	40	4	5	6	1	0	4				
23	05590	0	94630	48	1	3	10	1	0	4				
24	01504	0	94630	32	1	3	9	1	0	4				
25	06952	0	94670	56	3	2	10	1	0	4				
26	08089	0	94690	54	4	4	9	1	0	4				
27	03175	0	94800	55	2	1	6	1	0	4				
28	08891	0	94800	41	4	3	12	1	0	4				
29	05468	1	94810	39	4	2	10	1	0	4				
30	01767	0	94890	68	2	4	13	1	0	4				
31	09131	0	95000	53	1	2	8	1	0	4				
32	01913	0	95030	61	4	4	11	1	0	4				
33	05135	0	95080	61	1	3	7	1	0	4				
34	07533	0	95200	56	3	5	7	1	0	4				
35	02696	0	95270	55	4	5	9	1	0	4				

Data View Variable View

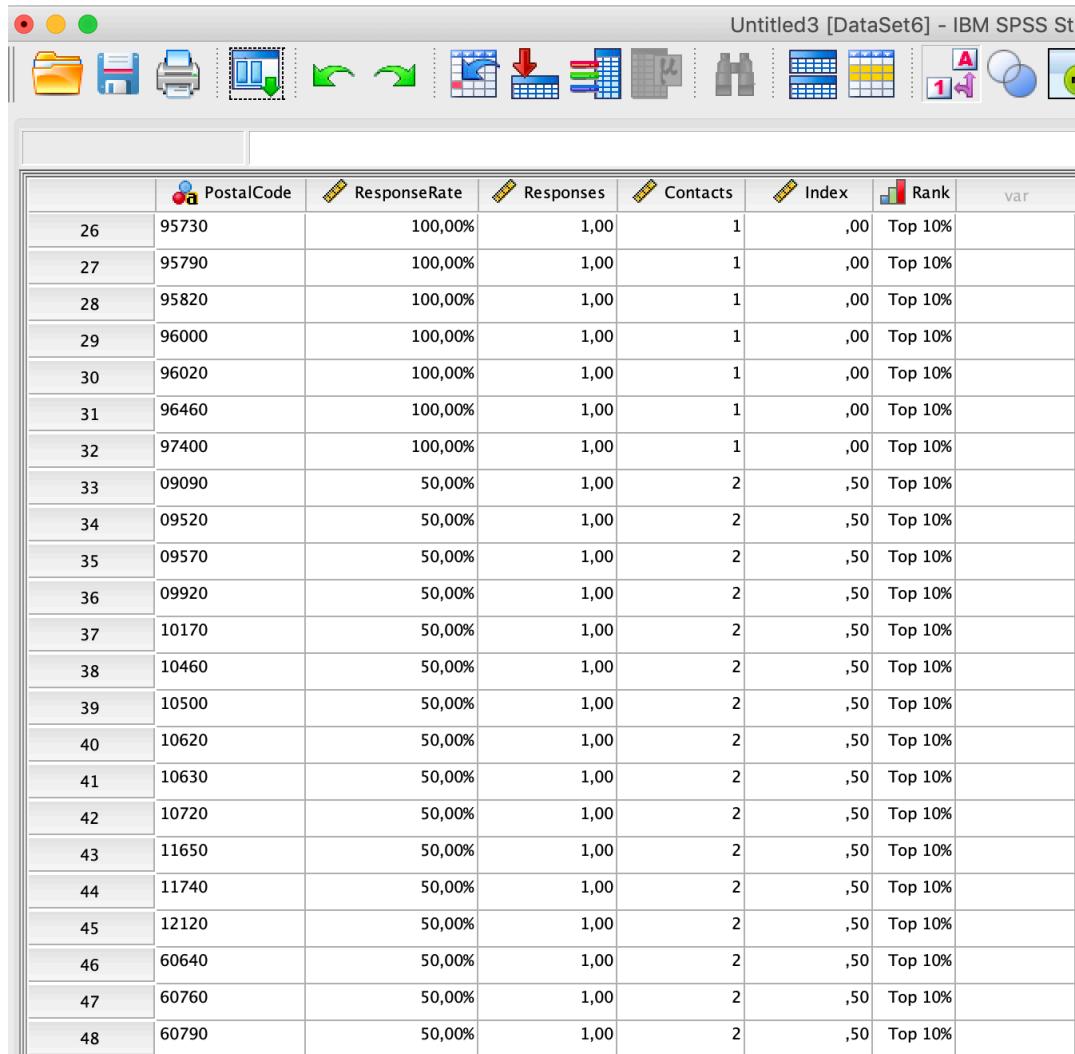
IBM SPSS Statistics Processor is ready Unicode:ON

# IBM® SPSS® Statistics

## Top postal codes



# IBM® SPSS® Statistics Results



	PostalCode	ResponseRate	Responses	Contacts	Index	Rank	var
26	95730	100,00%	1,00	1	,00	Top 10%	
27	95790	100,00%	1,00	1	,00	Top 10%	
28	95820	100,00%	1,00	1	,00	Top 10%	
29	96000	100,00%	1,00	1	,00	Top 10%	
30	96020	100,00%	1,00	1	,00	Top 10%	
31	96460	100,00%	1,00	1	,00	Top 10%	
32	97400	100,00%	1,00	1	,00	Top 10%	
33	09090	50,00%	1,00	2	,50	Top 10%	
34	09520	50,00%	1,00	2	,50	Top 10%	
35	09570	50,00%	1,00	2	,50	Top 10%	
36	09920	50,00%	1,00	2	,50	Top 10%	
37	10170	50,00%	1,00	2	,50	Top 10%	
38	10460	50,00%	1,00	2	,50	Top 10%	
39	10500	50,00%	1,00	2	,50	Top 10%	
40	10620	50,00%	1,00	2	,50	Top 10%	
41	10630	50,00%	1,00	2	,50	Top 10%	
42	10720	50,00%	1,00	2	,50	Top 10%	
43	11650	50,00%	1,00	2	,50	Top 10%	
44	11740	50,00%	1,00	2	,50	Top 10%	
45	12120	50,00%	1,00	2	,50	Top 10%	
46	60640	50,00%	1,00	2	,50	Top 10%	
47	60760	50,00%	1,00	2	,50	Top 10%	
48	60790	50,00%	1,00	2	,50	Top 10%	

# Thank You

**Yann Gouedo**

Data Scientist Leader – Machine Learning / Artificial Intelligence  
Marketing / Risk / Fraud / Maintenance  
IBM Certified Senior Data Scientist & IBM Certified Senior Architect

