

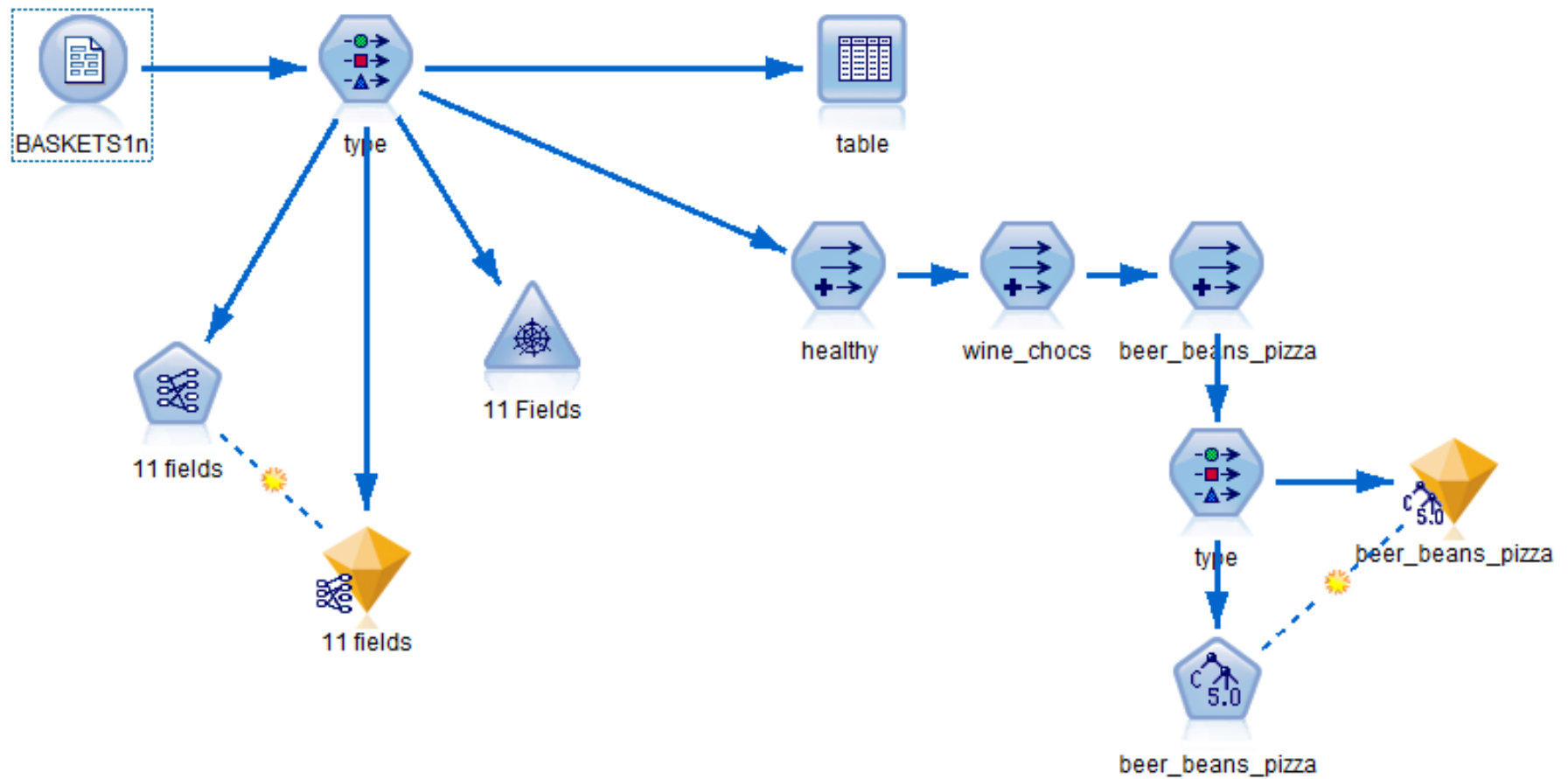
Association Analysis: A Case Study using SPSS Modeler

© 2016 NUS. The contents contained in this document may not be reproduced in any form or by any means, without the written permission of ISS, NUS, other than for the purpose for which it has been supplied.

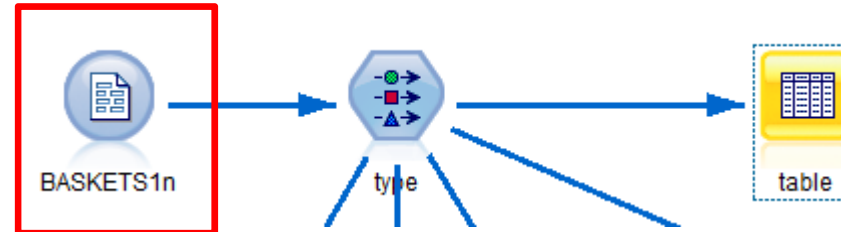
Context

- Supermarket transaction data
- With personal data of the purchaser, collected through the usage of loyalty cards
- Objective:
 - To discover groups of customers who buy similar products
 - To characterize (or profile) such groups demographically
- Using link analysis and association rule modeling to find products frequently bought together
- Using C5.0 rule induction to for group profiling

The Overall Stream



The Data File



BASKETS1n

Preview Refresh

\$CLEO_DEMOS/BASKETS1n

File Data Filter Types Annotations

File: \$CLEO_DEMOS/BASKETS1n

```
cardid,value,pmethod,sex,homeown,income,age,fruitveg,freshmeat,dairy,cannedveg,c
39808,42.7123,CHEQUE,M,NO,27000,46,F,T,T,F,F,F,F,F,F,T
67362,25.3567,CASH,F,NO,30000,28,F,T,F,F,F,F,F,F,F,T
10872,20.6176,CASH,M,NO,13200,36,F,F,F,T,F,T,T,F,F,T,F
```

☒ Read field names from file ☐ Specify number of fields

Skip header characters: 0 EOL comment characters:

Strip lead and trail spaces: ☒ None ☐ Left ☐ Right ☐ Both

Typing the Data

Diagram illustrating the data typing process:

A data source labeled **BASKETS1n** is connected to a **type** node (highlighted with a red box). The **type** node is connected to a **type** window.

The **type** window displays the following data types and roles:

Field	Measurement	Values	Missing	Check	Role
cardid	Typeless			None	None
value	Continuous	[10.007,49...		None	None
pmethod	Nominal	CARD,CA...		None	None
sex	Nominal	F,M		None	None
homeown	Flag	YES/NO		None	None
income	Continuous	[10200,30...		None	None
age	Continuous	[16,50]		None	None
fruitveg	Flag	T/F		None	Both
freshmeat	Flag	T/F		None	Both
dairy	Flag	T/F		None	Both
cannedveg	Flag	T/F		None	Both
cannedmeat	Flag	T/F		None	Both
frozenmeal	Flag	T/F		None	Both
beer	Flag	T/F		None	Both
wine	Flag	T/F		None	Both
softdrink	Flag	T/F		None	Both
fish	Flag	T/F		None	Both
confectionery	Flag	T/F		None	Both

The Variables

- **Basket summary:**
 - *cardid*. Loyalty card identifier for customer purchasing this basket.
 - *value*. Total purchase price of basket.
 - *pmethod*. Method of payment for basket.
- **Personal details of cardholder:**
 - *sex*
 - *homeown*. Whether or not cardholder is a homeowner.
 - *income*
 - *age*
- **Basket contents—flags for presence of product categories:**
 - *fruitveg*
 - *freshmeat*
 - *dairy*
 - *cannedveg*
 - *cannedmeat*
 - *frozenmeal*
 - *beer*
 - *wine*
 - *softdrink*
 - *fish*
 - *confectionery*

The Data Table

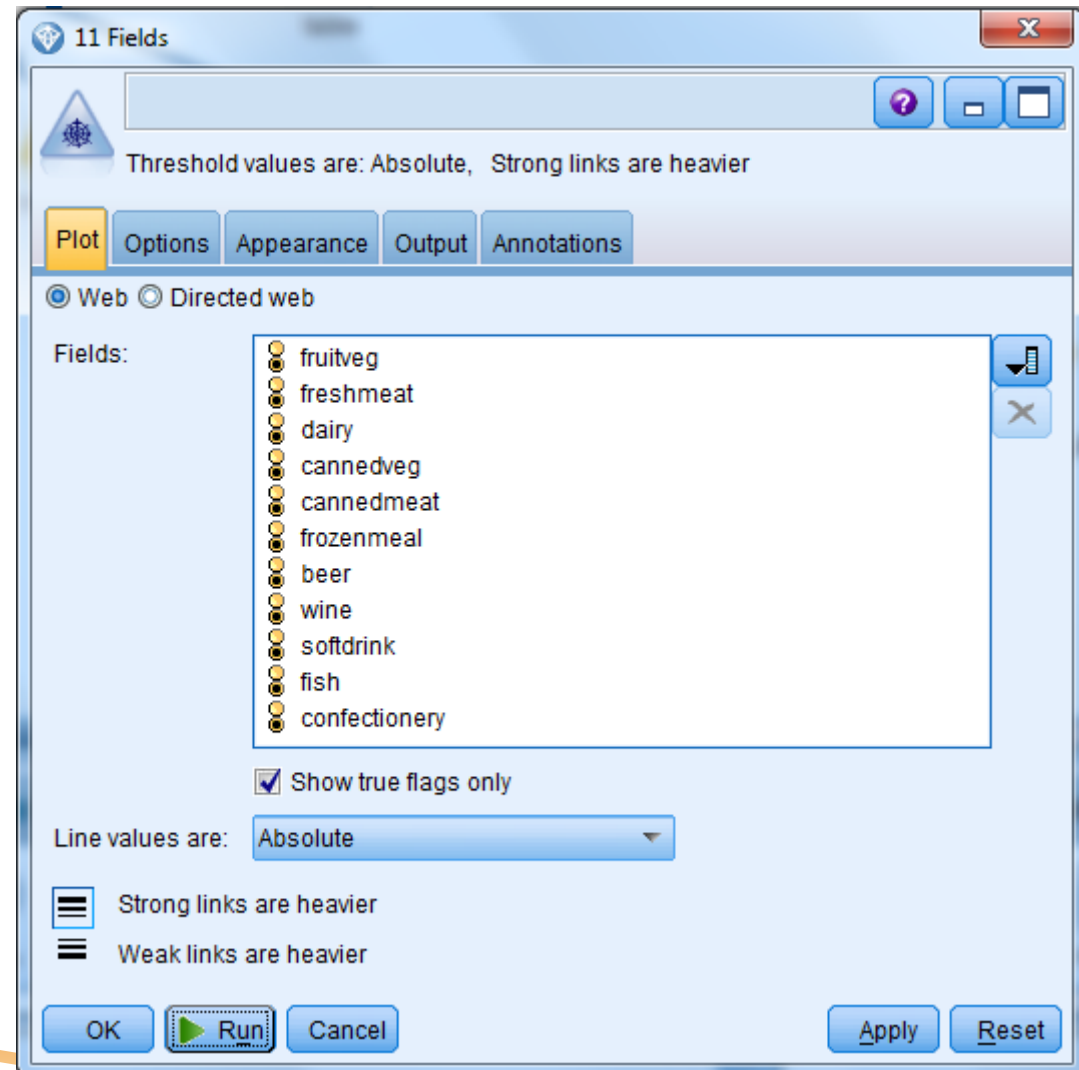
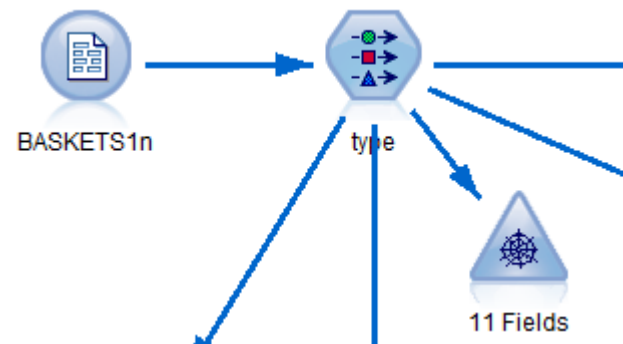
table (18 fields, 1,000 records) #1

File Edit Generate

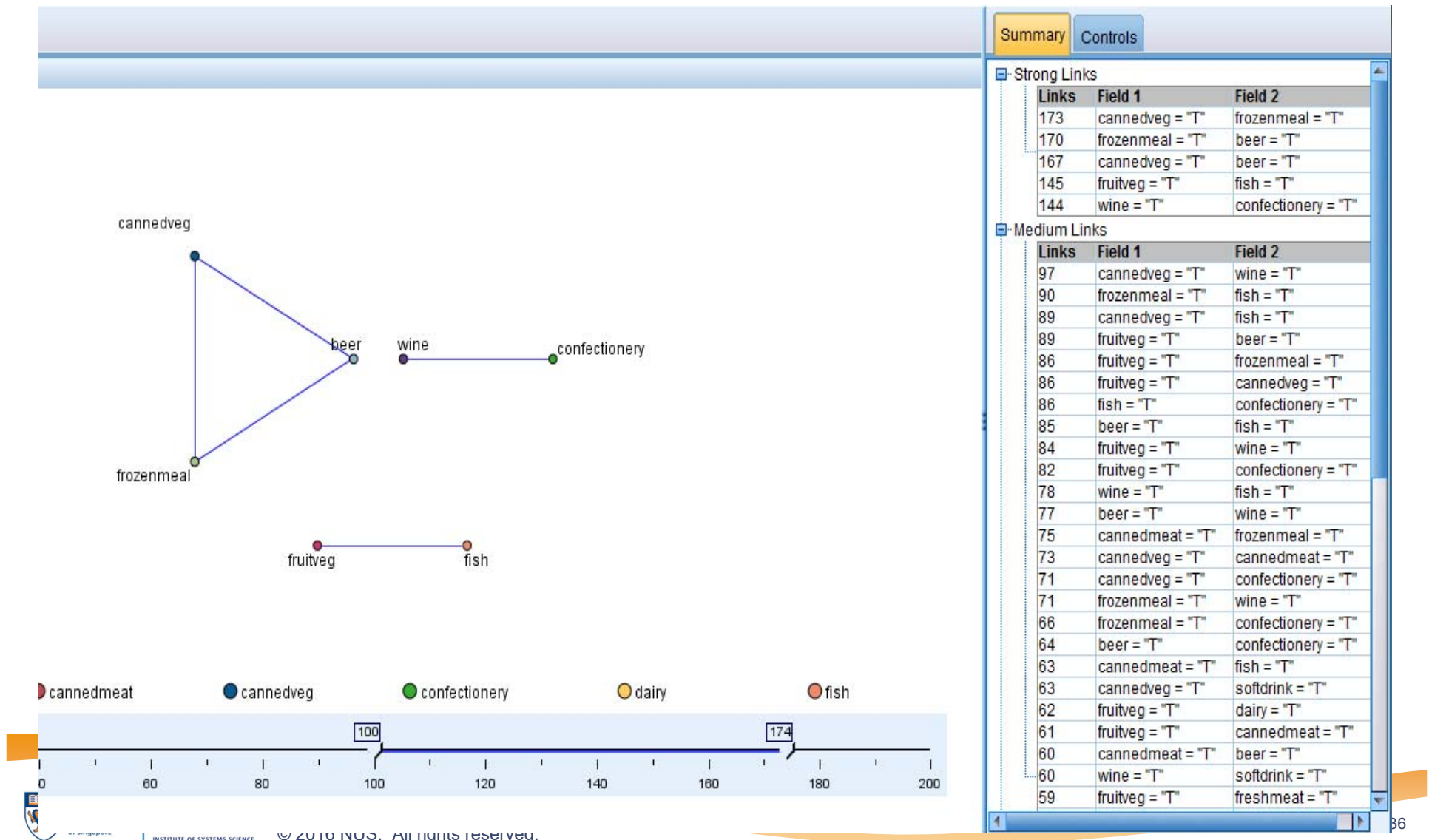
Table Annotations

	cardid	value	pmethod	sex	homeown	income	age	fruitveg	freshmeat	dairy	cannedveg	cannedmeat	frozenmeal
1	39808	42.712	CHEQUE	M	NO	27000	46 F	T	T	F	F	F	F
2	67362	25.357	CASH	F	NO	30000	28 F	T	F	F	F	F	F
3	10872	20.618	CASH	M	NO	13200	36 F	F	F	T	F	F	T
4	26748	23.688	CARD	F	NO	12200	26 F	F	T	F	F	F	F
5	91609	18.813	CARD	M	YES	11000	24 F	F	F	F	F	F	F
6	26630	46.487	CARD	F	NO	15000	35 F	T	F	F	F	F	F
7	62995	14.047	CASH	F	YES	20800	30 T	F	F	F	F	F	F
8	38765	22.203	CASH	M	YES	24400	22 F	F	F	F	F	F	F
9	28935	22.975	CHEQUE	F	NO	29500	46 T	F	F	F	F	F	T
10	41792	14.569	CASH	M	NO	29600	22 T	F	F	F	F	F	F
11	59480	10.328	CASH	F	NO	27100	18 T	T	T	T	F	F	F
12	60755	13.780	CASH	F	YES	20000	48 T	F	F	F	F	F	F
13	70998	36.509	CARD	M	YES	27300	43 F	F	T	F	T	T	T
14	80617	10.201	CHEQUE	F	YES	28000	43 F	F	F	F	F	F	F
15	61144	10.374	CASH	F	NO	27400	24 T	F	T	F	F	F	F
16	36405	34.822	CHEQUE	F	YES	18400	19 F	F	F	F	F	F	T
17	76567	42.248	CARD	M	YES	23100	31 T	F	F	T	F	F	F
18	85699	18.169	CASH	F	YES	27000	29 F	F	F	F	F	F	F
19	11357	10.753	CASH	F	YES	23100	26 F	F	F	F	F	F	F
20	97761	32.318	CARD	F	YES	25800	38 T	F	F	T	F	F	F
21	20362	31.720	CASH	M	YES	25100	38 F	F	F	F	F	F	T

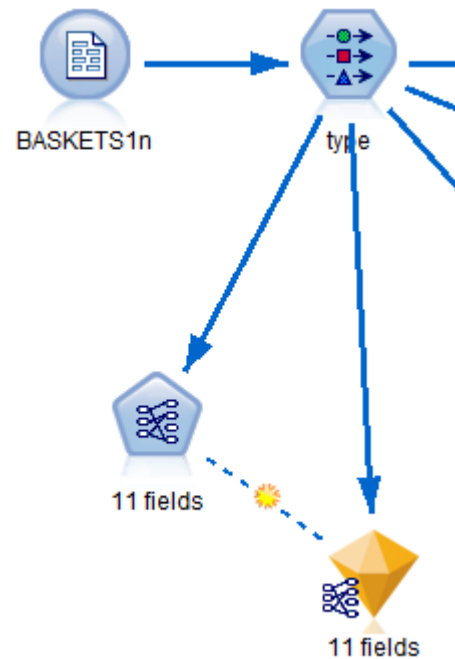
Link Analysis



Revealing 3 Groups



Association Rules Modelling



11 fields

Fields Model Expert Annotations

Model name: ☒ Auto ☐ Custom

☒ Use partitioned data

Minimum antecedent support (%):

Minimum rule confidence (%):

Maximum number of antecedents:

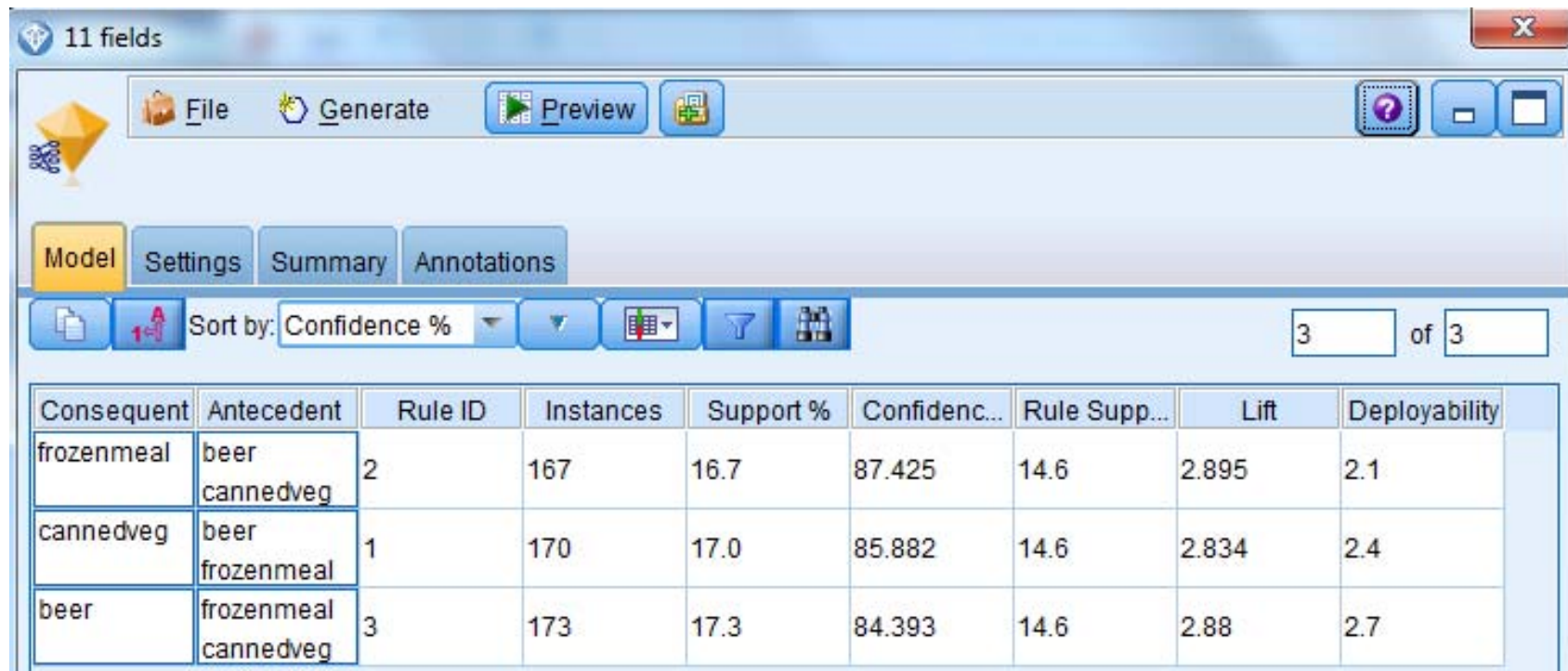
☒ Only true values for flags

Optimize: ☒ Speed ☐ Memory

OK Run Cancel Apply Reset

Rules Found

- Based on the given minimum antecedent support and minimum confidence



11 fields

File Generate Preview

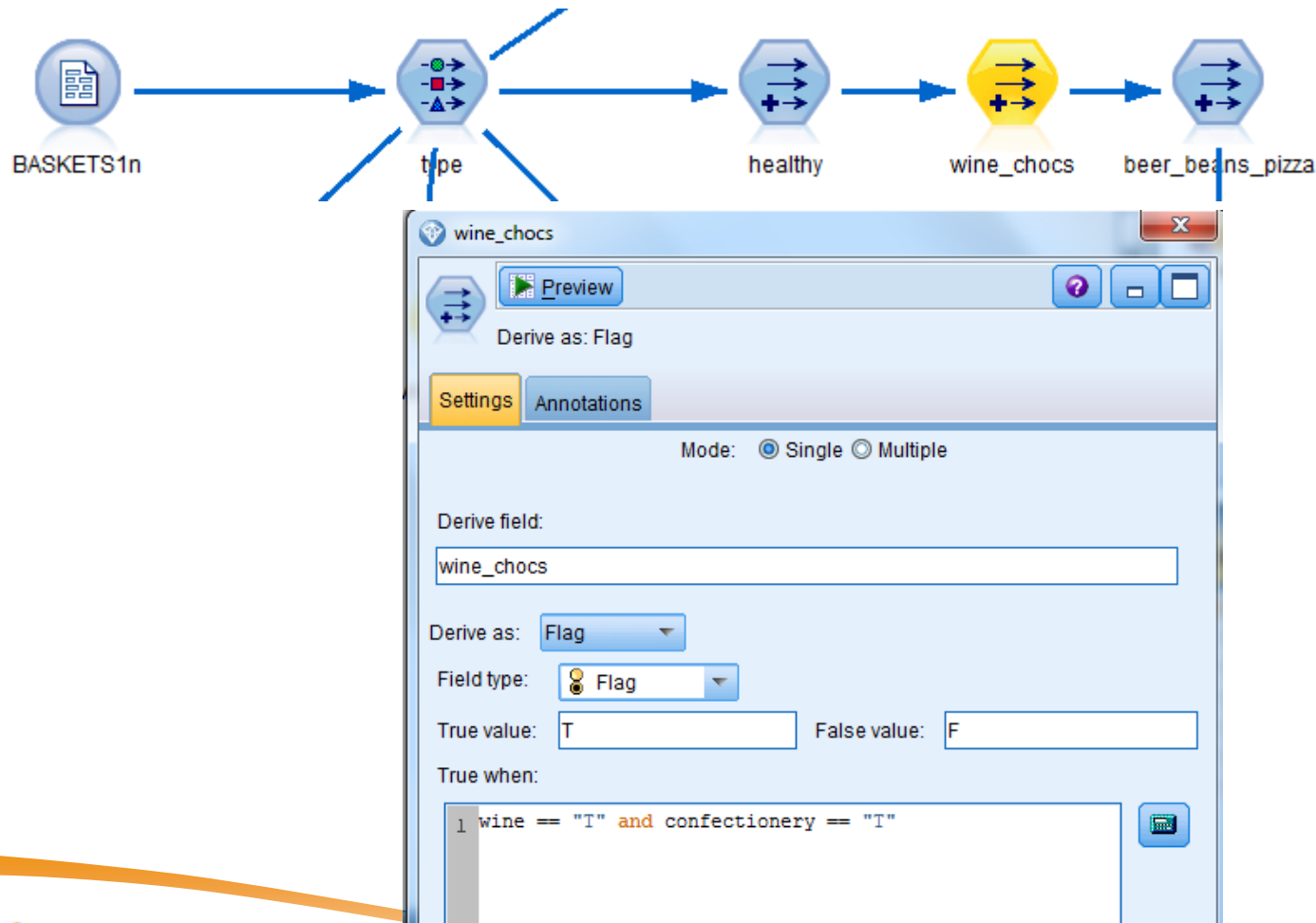
Model Settings Summary Annotations

Sort by: Confidence % 3 of 3

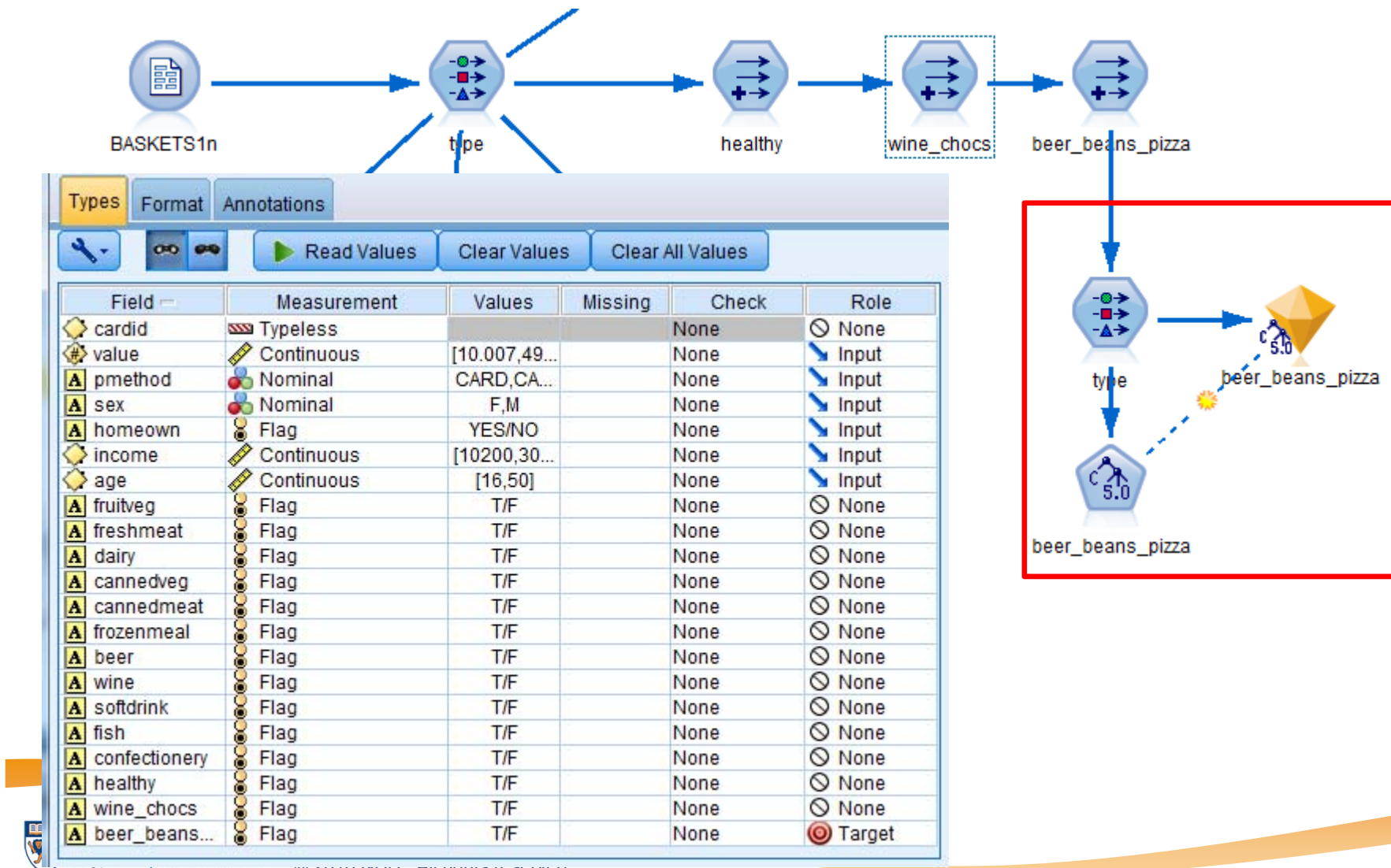
Consequent	Antecedent	Rule ID	Instances	Support %	Confidenc...	Rule Supp...	Lift	Deployability
frozenmeal	beer cannedveg	2	167	16.7	87.425	14.6	2.895	2.1
cannedveg	beer frozenmeal	1	170	17.0	85.882	14.6	2.834	2.4
beer	frozenmeal cannedveg	3	173	17.3	84.393	14.6	2.88	2.7

Deriving Group Variables

- Derive one variable for each group



Rule Induction for Profiling



Rule Induction

- Information about this group of customers!

beer_beans_pizza

5.0

Fields Model Costs Analyze Annotations

Model name: ☒ Auto ☐ Custom beer_beans_pizza

☒ Use partitioned data

☒ Build model for each split

Output type: ☐ Decision tree ☒ Rule set

☐ Group symbolics

☐ Use boosting Number of trials: 10

☐ Cross-validate Number of folds: 10

Mode: ☒ Simple ☐ Expert

Favor: ☒ Accuracy ☐ Generality

Expected noise (%): 0

Rules for T - contains 1 rule(s)

Rule 1 for T

if sex = M
and income <= 16,900
then T

Rules for F - contains 2 rule(s)

Rule 1 for F

if income > 16,900
then F

Rule 2 for F

if sex = F
then F

Default: F

Homework (unmarked)

- What about the other two groups?
- Try your hands to find out.