SCHOOL OF COMPUTER APPLICATION

Case Study On

ACME, a company selling sports products, wants to promote its new product: the XL Original Orange Baseball Cap. To test customer interest, ACME sent a test mailing to 10,000 randomly selected customers and recorded their responses.
Two datasets are provided:

Submitted To:
Mr Ayushman Bhadauriya

Submitted By:
Srijan Jaiswal

Babu Banarasi Das University, Lucknow

BBD city, Faizabad road, Lucknow Uttarpradesh- 226 028 India

ACME Case Study: Predicting Customer Response

ACME, a company selling sports products, wants to promote its new product: the XL Original Orange Baseball Cap. To test customer interest, ACME sent a test mailing to 10,000 randomly selected customers and recorded their responses. Two datasets are provided:

train_acme_customers.sav – training dataset containing customers who received the test mailing and their responses.

test_acme_customers.sav – testing dataset containing customers who did not receive the test mailing (response field undefined).

Both datasets include the following fields:

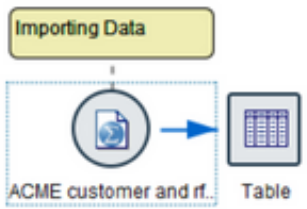
- **customer_id – customer's identification number**
- **gender – customer's gender**
- **email_address – customer's e-mail address**
- **postal_code – customer's postal code**
- **recency_01_01_2011 – last order date before Jan 1, 2011 frequency_01_01_2011 – number of orders before Jan 1, 2011 monetary_value_01_01_2011 – total purchase amount before Jan 1, 2011 has_received_test_mailing – flag whether the customer received the Feb 1, 2011 test mailing**
- **response – whether the customer ordered the XL Original Orange Baseball Cap (only valid for training dataset customers)**
- **orderdate – date the cap was ordered (only for respondents) days_between_test_and_order – days between test mailing and orderdate (only for respondents)**
- **ordered_within_month – whether the order happened within one month after mailing (only for respondents)**

- **Required Tool:** IBM SPSS Modeler

Task-1

1. Import the training dataset into IBM SPSS Modeler-

From Source node Add Statistics File and Connect a Table to show the Data.



2.Run a Table node to summarize the data.

Right click on table and run it

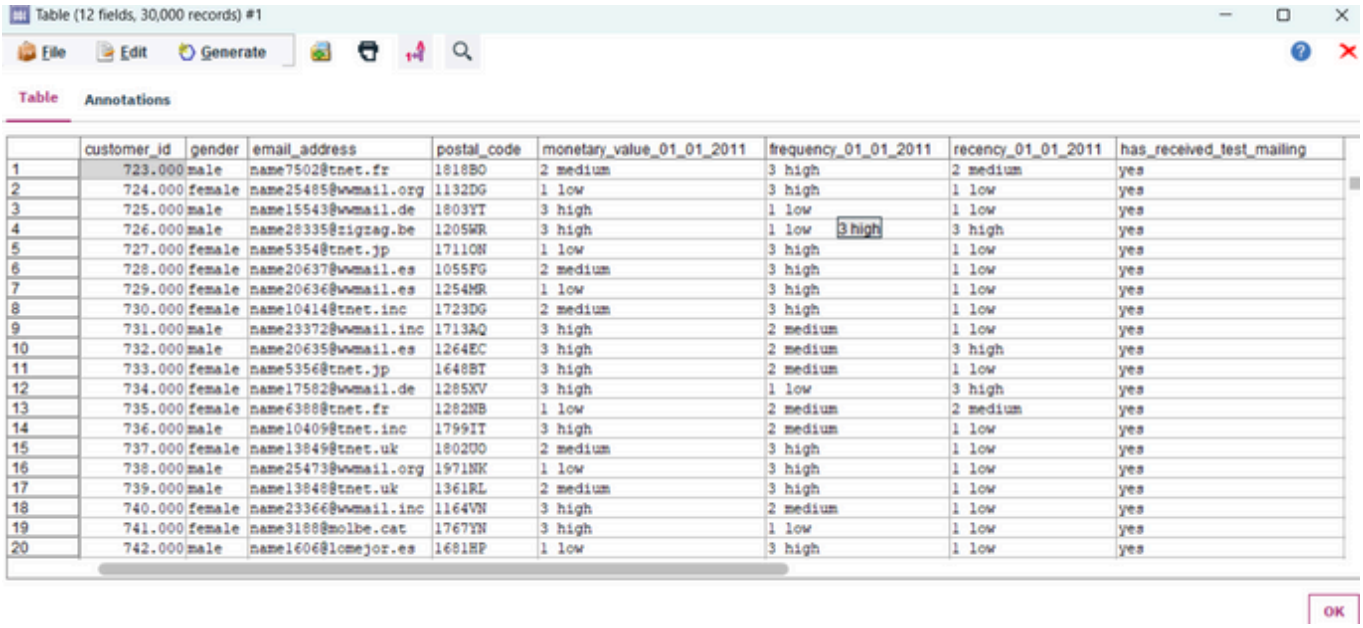


Table (12 fields, 30,000 records) #1

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
1	723.000	male	name7502@tnet.fr	1818BO	2 medium	3 high	2 medium	yes
2	724.000	female	name25485@wwmail.org	1132DG	1 low	3 high	1 low	yes
3	725.000	male	name15543@wwmail.de	1803YT	3 high	1 low	1 low	yes
4	726.000	male	name28335@zigzag.be	1205WR	3 high	1 low 3 high	3 high	yes
5	727.000	female	name5354@tnet.jp	1711ON	1 low	3 high	1 low	yes
6	728.000	female	name20637@wwmail.es	1055FG	2 medium	3 high	1 low	yes
7	729.000	female	name20636@wwmail.es	1254MR	1 low	3 high	1 low	yes
8	730.000	female	name10414@tnet.inc	1723DG	2 medium	3 high	1 low	yes
9	731.000	male	name23372@wwmail.inc	1713AQ	3 high	2 medium	1 low	yes
10	732.000	male	name20635@wwmail.es	1264EC	3 high	2 medium	3 high	yes
11	733.000	female	name5356@tnet.jp	1648BT	3 high	2 medium	1 low	yes
12	734.000	female	name17582@wwmail.de	1285XV	3 high	1 low	3 high	yes
13	735.000	female	name6388@tnet.fr	1282NB	1 low	2 medium	2 medium	yes
14	736.000	male	name10409@tnet.inc	1799IT	3 high	2 medium	1 low	yes
15	737.000	female	name13849@tnet.uk	1802DO	2 medium	3 high	1 low	yes
16	738.000	male	name25473@wwmail.org	1971NK	1 low	3 high	1 low	yes
17	739.000	male	name13848@tnet.uk	1361RL	2 medium	3 high	1 low	yes
18	740.000	female	name23366@wwmail.inc	1164VN	3 high	2 medium	1 low	yes
19	741.000	female	name3188@molbe.cat	1767YN	3 high	1 low	1 low	yes
20	742.000	male	name1606@lomejor.es	1681HP	1 low	3 high	1 low	yes

OK

3.How many records are in the training dataset?

30000 Records in Data.

4. How many fields are in the training dataset?

12 Fields in Data

Task-2

1. Select only customers who were in the test mailing

Run the table and in (has_recieved_text_mailing) column Select Yes Data ,then on the top there is a Generate Option click and select And node

1. How many customers were included in the test mailing?

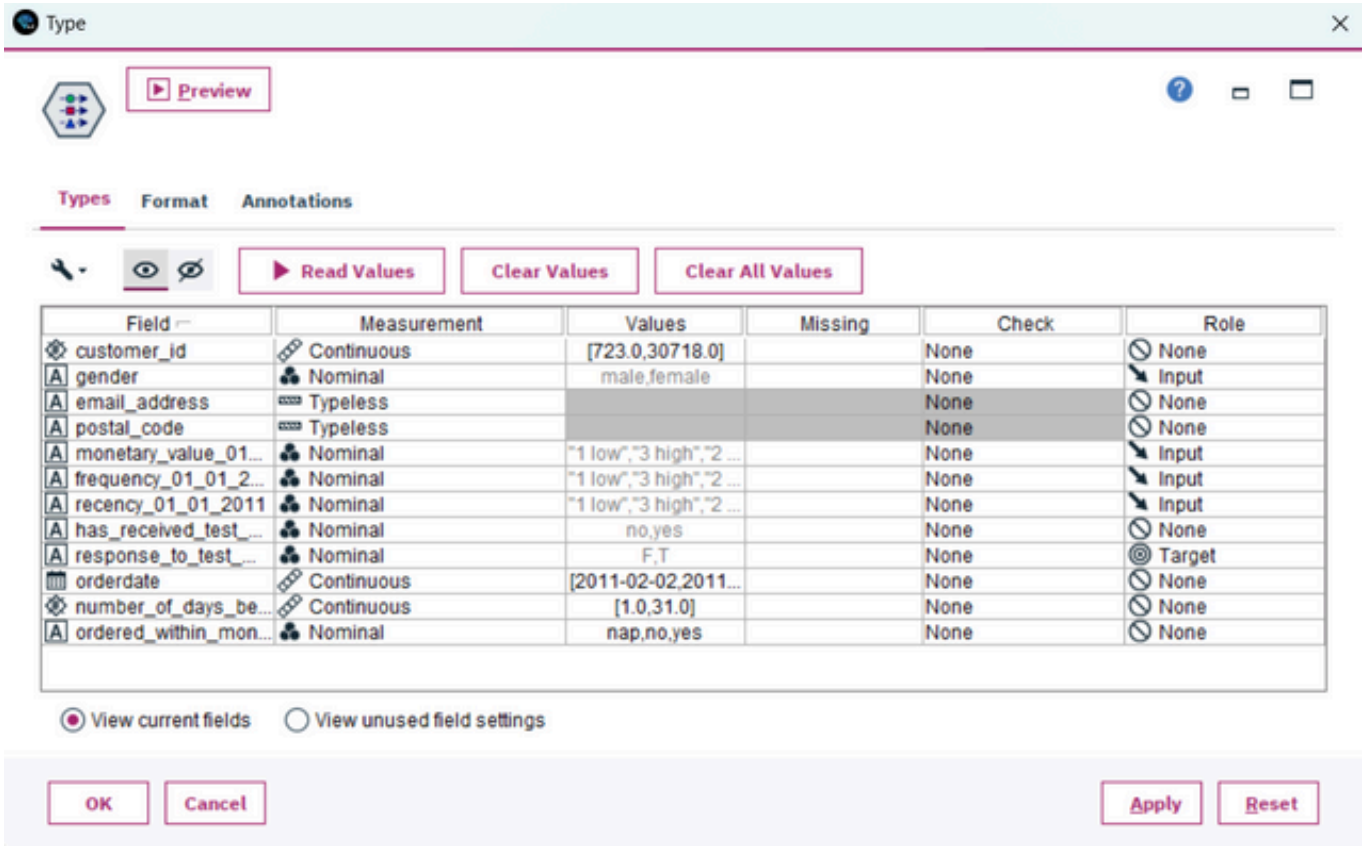
12 Fields and 10000 Records Included in test mailing

Task-3

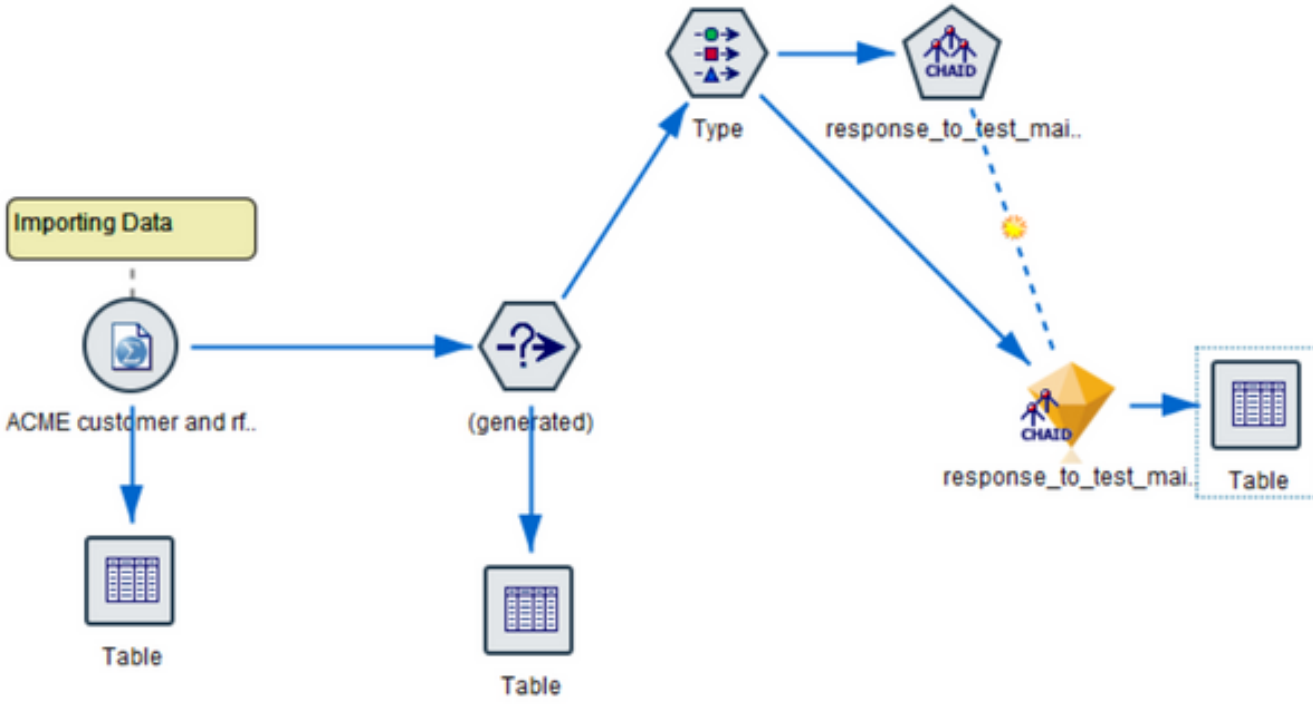
1.Build a CHAID decision tree model to predict response , using:

- gender
- recency_01_01_2011
- frequency_01_01_2011
- monetary_value_01_01_2011

Add Type Node form (Filed Ops),then select columns,Set the Target Column and rest of the column will be None.



Add Chaid model from Modeling node then run the model After this connect a table to the model to see output.



1. Which field is used as the first split?

monetary_value_01_01_2011 use as a first Spilt.

1. Which group shows the highest response rate? What is the probability of responding for this group?

Task-4

1.Run a Table node downstream of the model nugget.

Right click on the model and run it.

2. Identify the two new fields added by the model.

two columns added by model

- $R-response_to_test_mailing_02_01_2011
- $RC- response_to_test_mailing_02_01_2011

1. What do these fields represent?

first Column $R-response_to_test_mailing_02_01_2011 – Represents the Flase and True response by Customers

Second Column $RC-response_to_test_mailing_02_01_2011 – Represents the Probability that tell the Customer my buy Goods

$R-response_to_test_mailing_02_01_2011	$RC-response_to_test_mailing_02_01_2011
F	0.953
F	0.993
F	0.953
F	0.953
F	0.993
F	0.953
F	0.993
F	0.953
F	0.911
T	0.625
F	0.911
F	0.953
F	0.998
F	0.911
F	0.953
F	0.993
F	0.953
F	0.911
F	0.953
F	0.993

Task-5

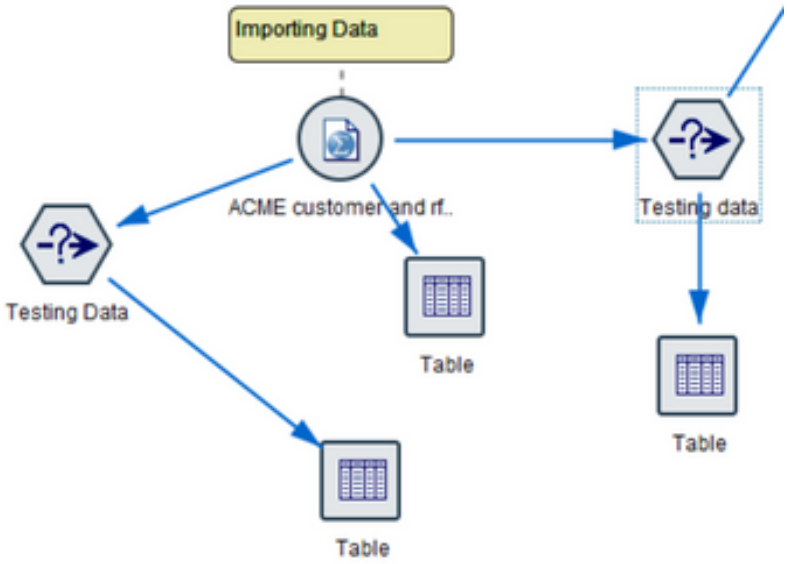
1. Apply the model to the testing dataset (customers who did not receive the test mailing).

1.Select the data of customers that did not receive mail

2.then go to Generate and Select And to create Generate the Data

3.connect a table to this node

4.then copy the Prevoius model and paste it on the side of Testing Data Node and connect Them ,Connect Table to see Result.



2. How many customers are predicted to respond positively (predicted = T)?

Click on the Output Table of Testing Data and Select the T Response and then Generate a node .

and Connect the node with the model that we Copied and connect to the Testing Data.

14 Fields and 254 Records Find Positive.

Task-6

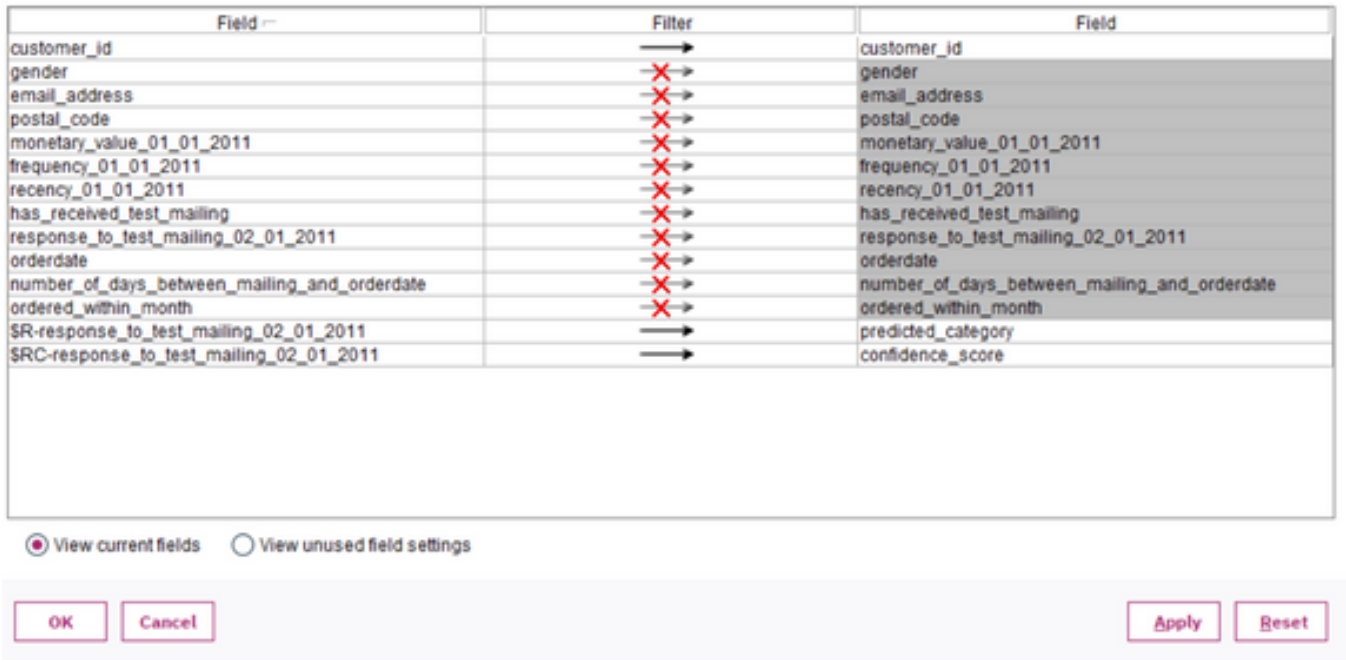
1. Export the selected customers to a text file name customers_to_contact.txt **Include only the following fields:**

customer_id

predicted category (rename to predicted_category)

confidence score (rename to confidence_score)

Connect a Filter Node From (Field Ops) and Remove the Node that not Needed.



From Export Node Add Flat File And Connect to the Filter Node and save the File Where you Want.