

**🡪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).**

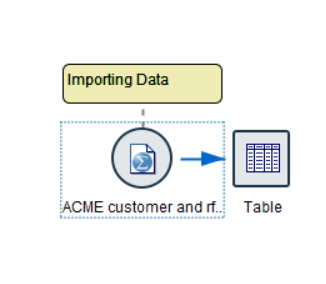
**Both datasets include the following fields:**

* **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**

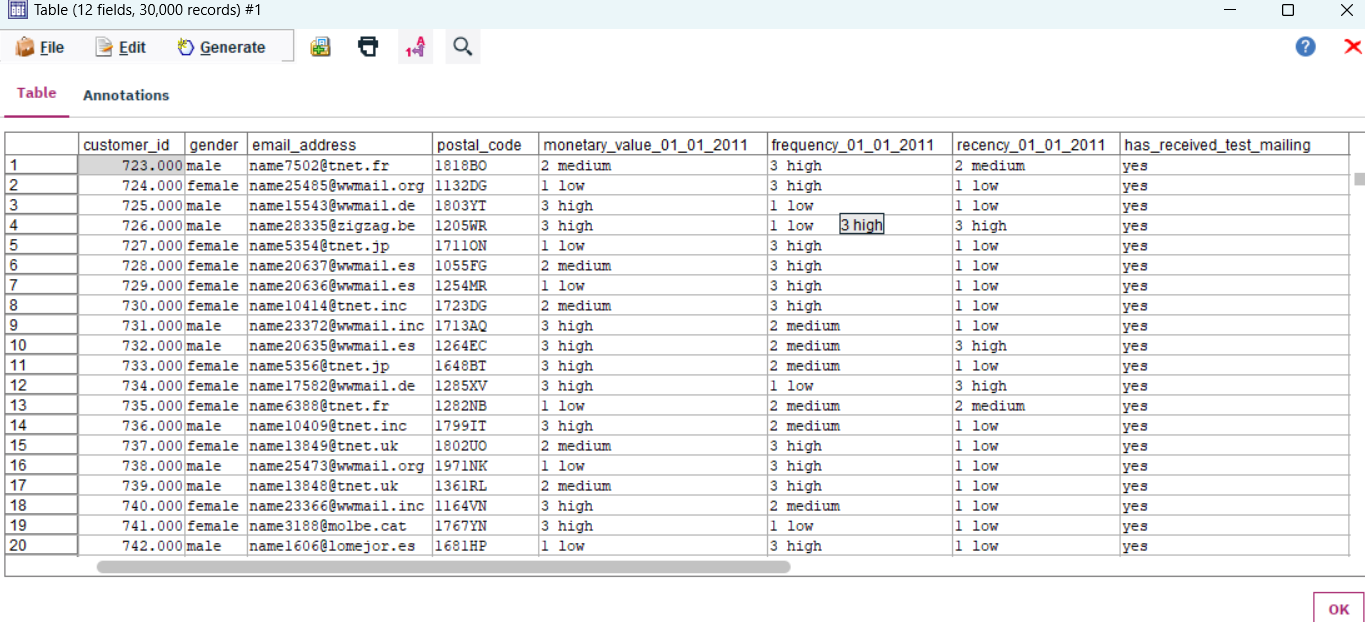
1. Import the training dataset into IBM SPSS Modeler-

🡪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



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?

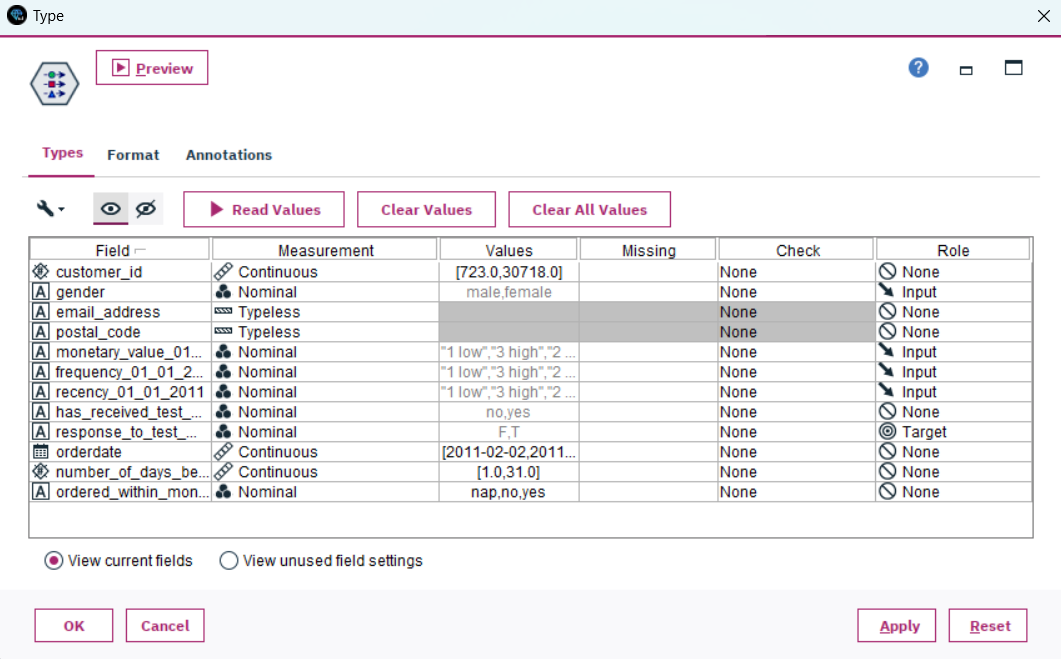
🡪12 Fields and 10000 Records Included in test mailing

**Task-3**

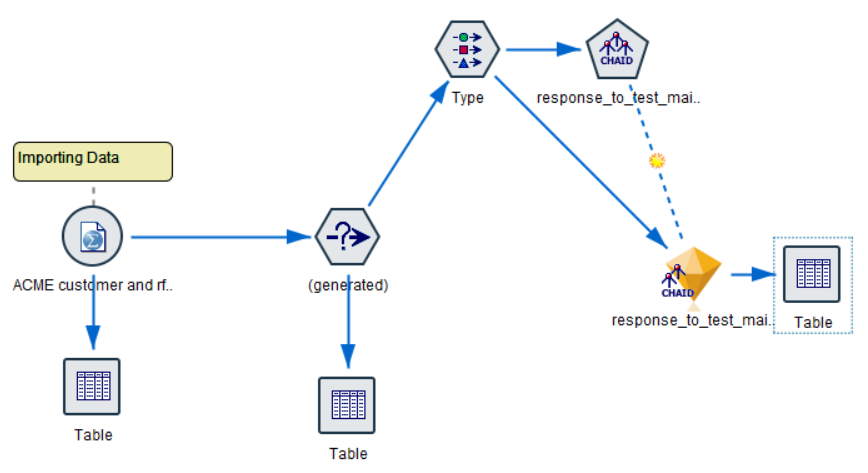
1.Build a CHAID decision tree model to predict response , using:

* 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.

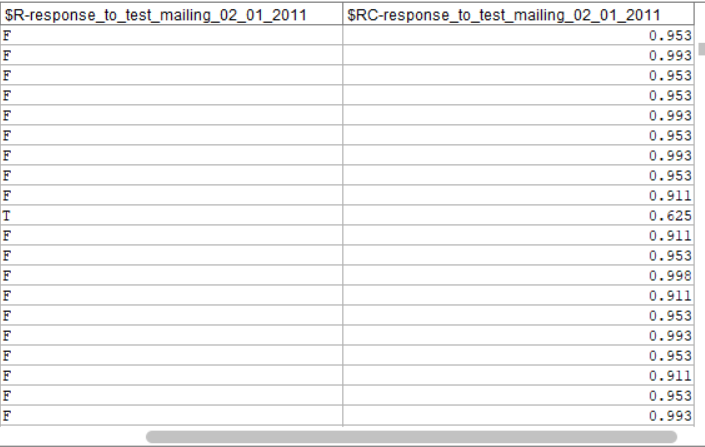
🡪 two columns added by model

* + - $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



**Task-5**

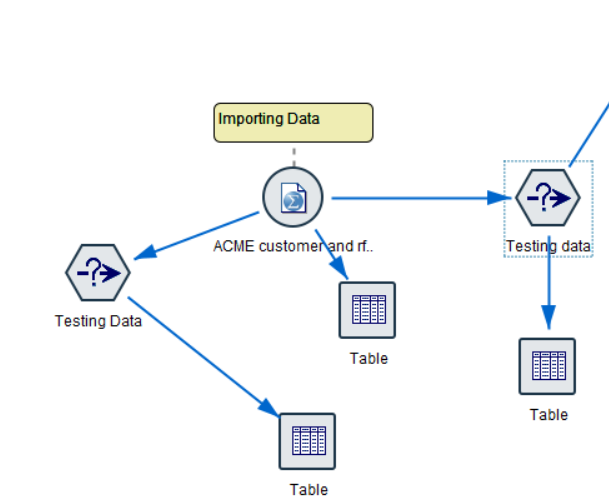
1. Apply the model to the testing dataset (customers who did not receive the test mailing).

🡪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**

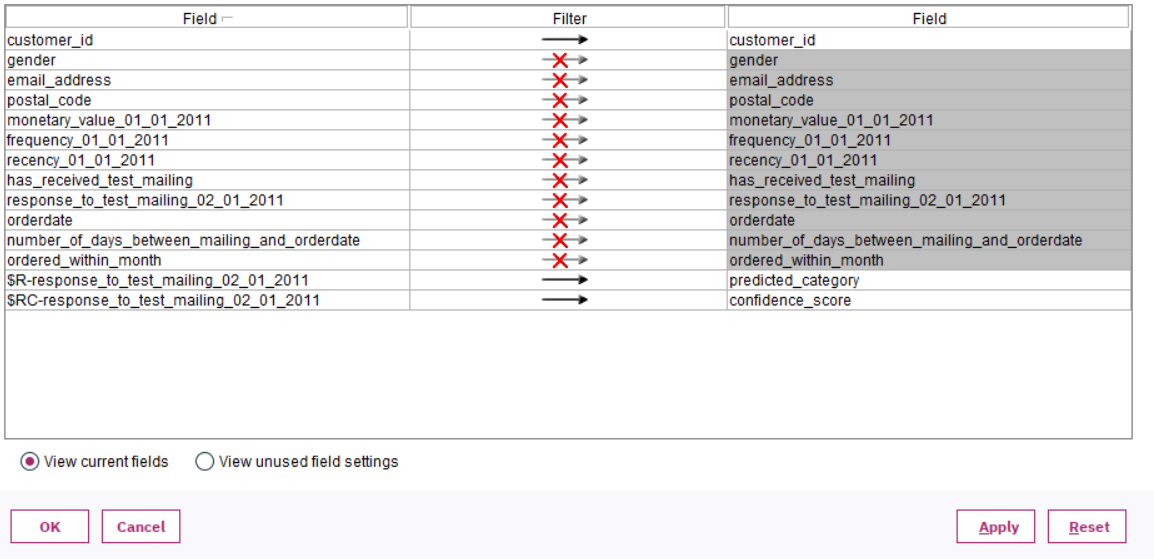
1. Export the selected customers to a text file name customers\_to\_contact.txt **Include only the following fields:**

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.

