

SCHOOL OF COMPUTER APPLICATION

PREDICTIVE ANALYTICS

BCADS1501



Submitted By:

Name: Siddhant Kumar Patel

Section: BCA DS 36

University Roll No: 1230258423

Semester: 5th

Submitted To:

Mr. Ayushman Bhadauria

Babu Banarasi Das University, Lucknow

BBD City, Ayodhya Road, Lucknow Uttar Pradesh- 226028 Bharat

Index

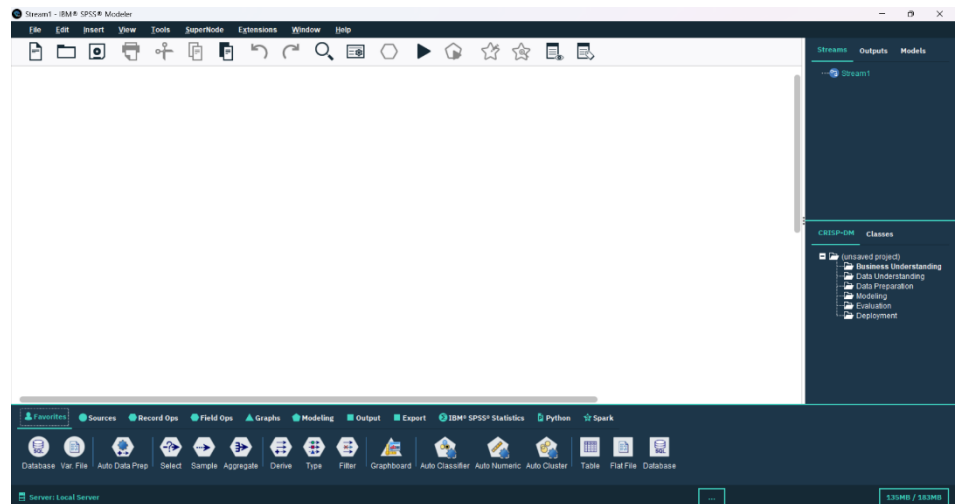
[illegible]

Practical

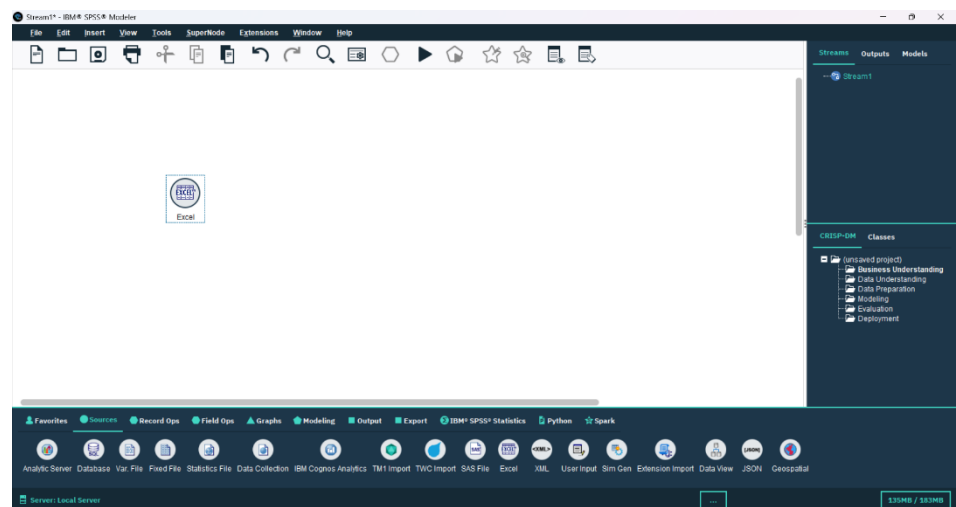
1. Import the dataset to the SPSS modular and run it.

Steps:

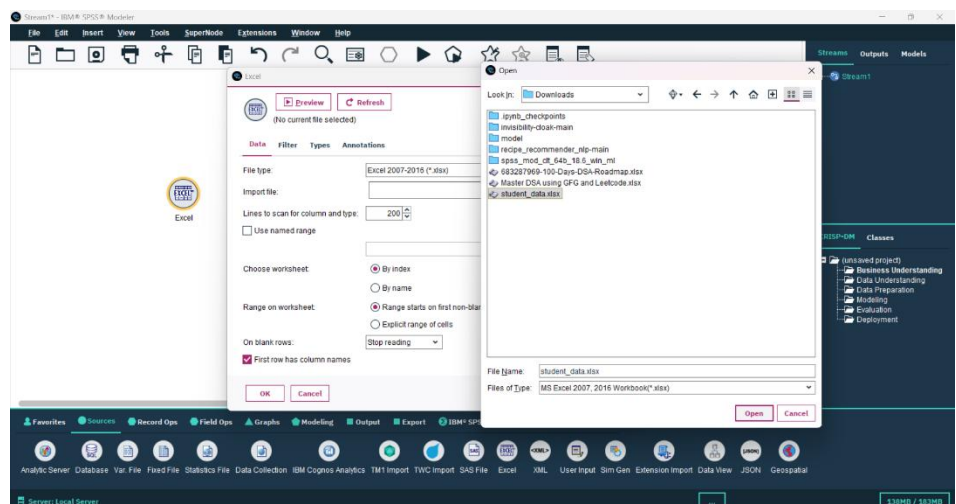
1. Open the SPSS Modular



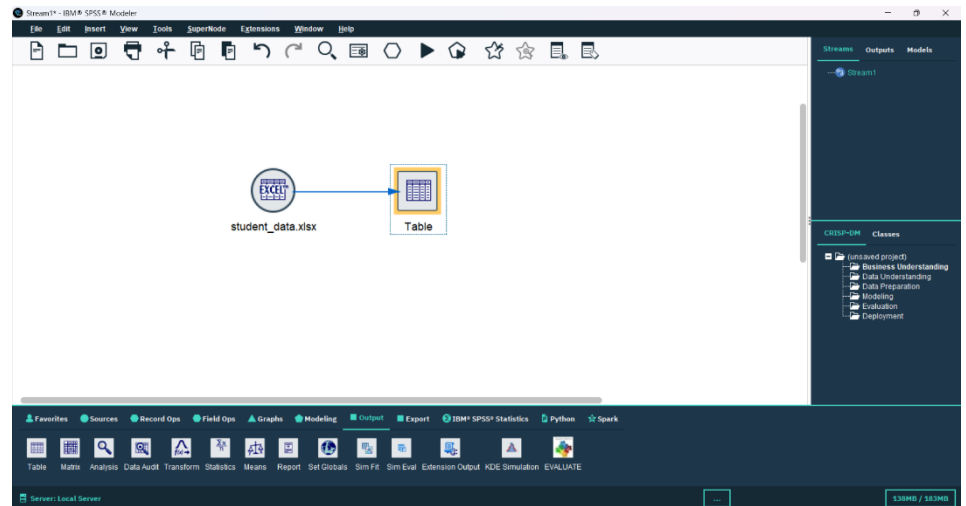
2. Click on source option, drag and drop the Excel node to import the excel file



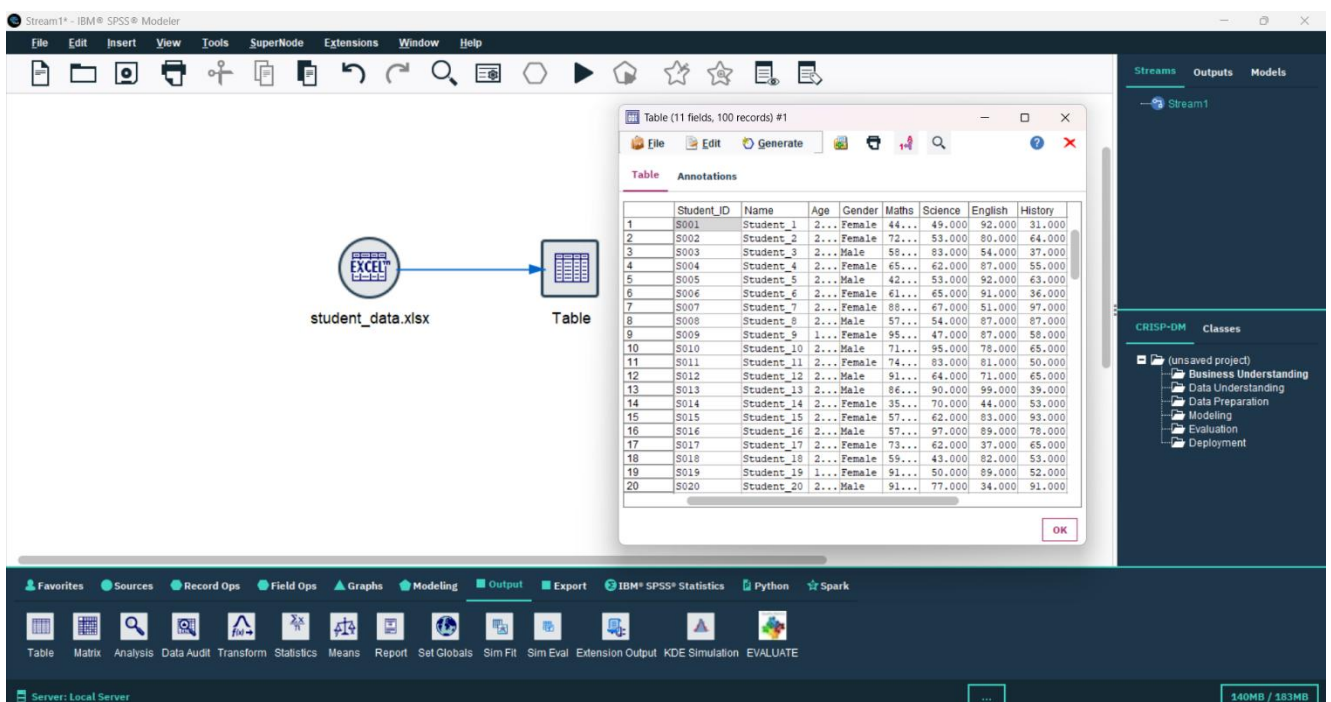
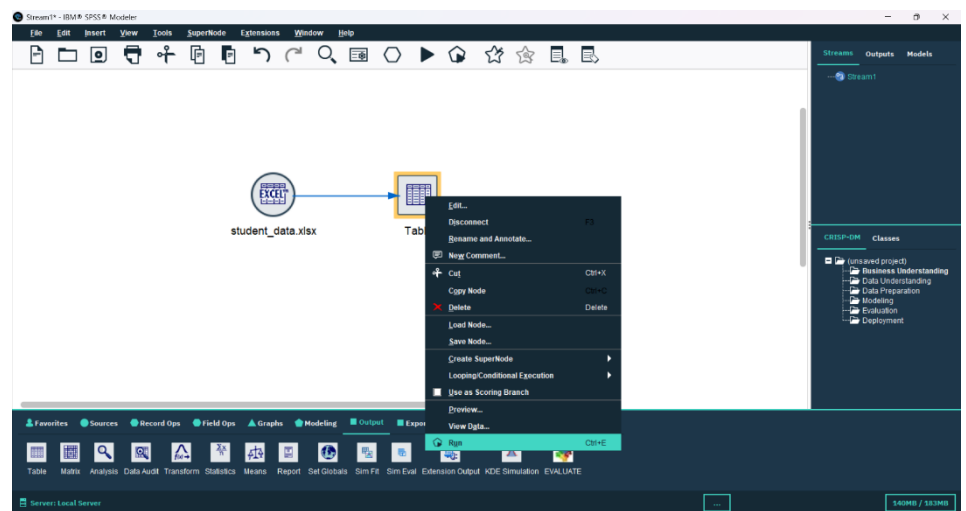
3. Click on the excel node to import the data from the file path



- Go to output, drag and drop the table node and connect with the excel node



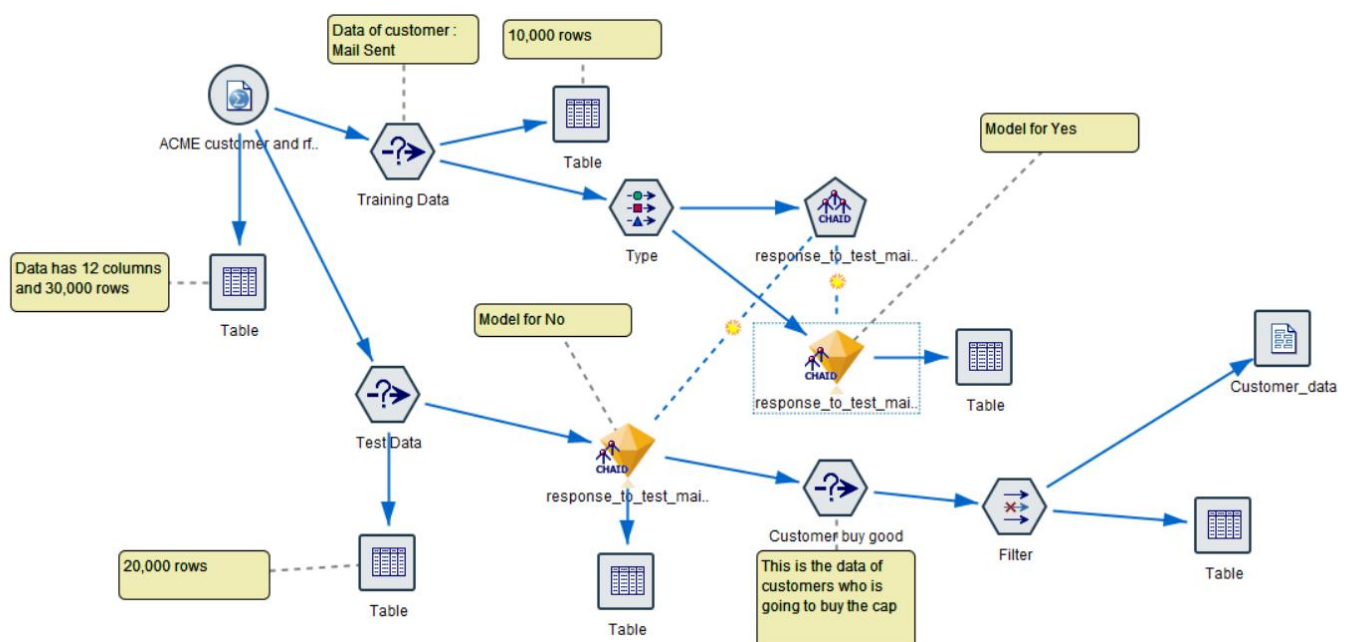
- Run the table node to see the result



2. Use the ACME dataset, make a model and predict the output.

Steps:

1. Import the dataset, use the table to get all the data (12 columns and 30000 rows)
2. Filter the data for the "YES" value and generate the node, rename as training data.
3. Connect that node with table to get the desired output. (10000 rows)
4. Drag and drop the type node and connect with the training data node.
5. Use the CHAID model and train the model with data.
6. Use the table to get the result after the model is trained.
7. Go back and filter the data for the "NO" value and generate the node, rename as test data for testing out model.
8. Make the copy of the model and connect with the test data.
9. Use the table along with all nodes to get the result.
10. Again generate a new node for the "TRUE" data in the model. (Data of customers who are going to buy the cap)
11. Use the filter node to rename the specific columns and filter the data that we needed.
12. After all export the data to your local file with choosing the flat file node and specify the path where you want to save.



Q 1. How many records are in the training dataset?

Ans. There are 30000 Rows in the dataset.

Q 2. How many fields are in the training dataset?

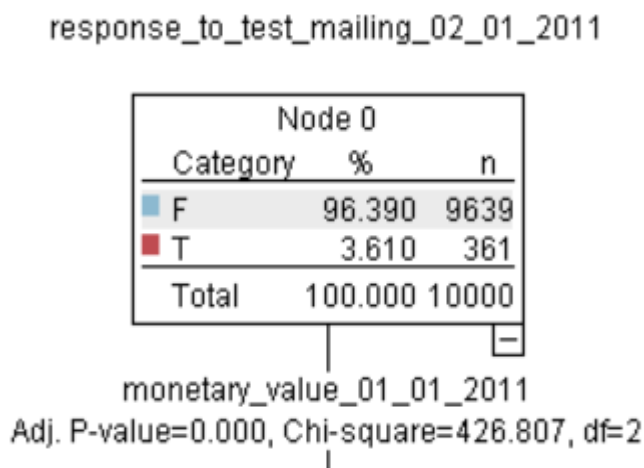
Ans. There are 12 Columns in the dataset.

Q 3. How many customers were included in the test mailing?

Ans. There are 20000 customers are included in the test mailing.

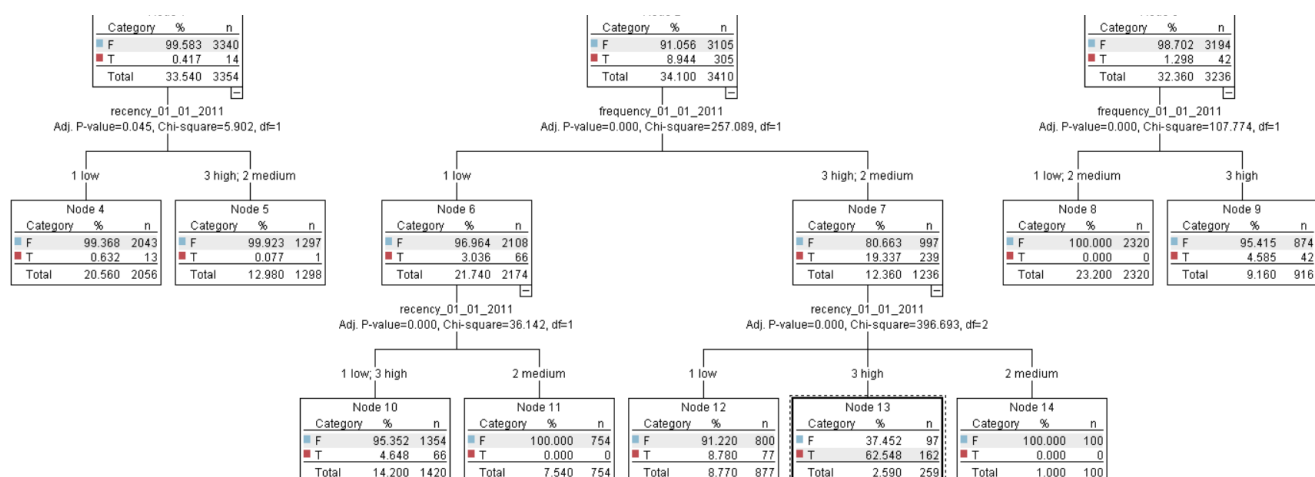
Q 4. Which field is used as the first split?

Ans. monetary_value_01_01_2011 fields is used as the first split.



Q 5. Which group shows the highest response rate? What is the probability of responding for this group?

Ans. recency_01_01_2011 shows the highest response rate and the probability of responding for this group is 62.548 %



Q 6. Identify the two new fields added by the model.

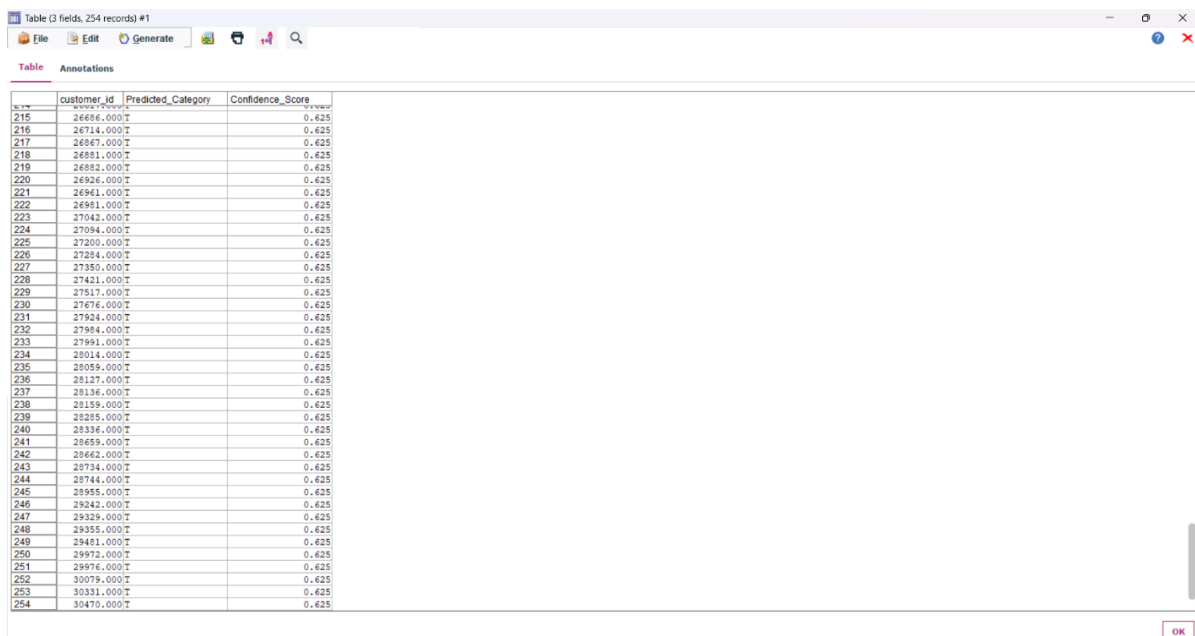
Ans. \$R-response_to_test_mailing_02_01_2011 and \$RC-response_to_test_mailing_02_01_2011 are two new fields added by the model.

Q 7. What do these fields represent?

Ans. The first fields represents the True/False value for the customers who are going to buy the cap or not. The second fields represents the probability of that condition is going to true.

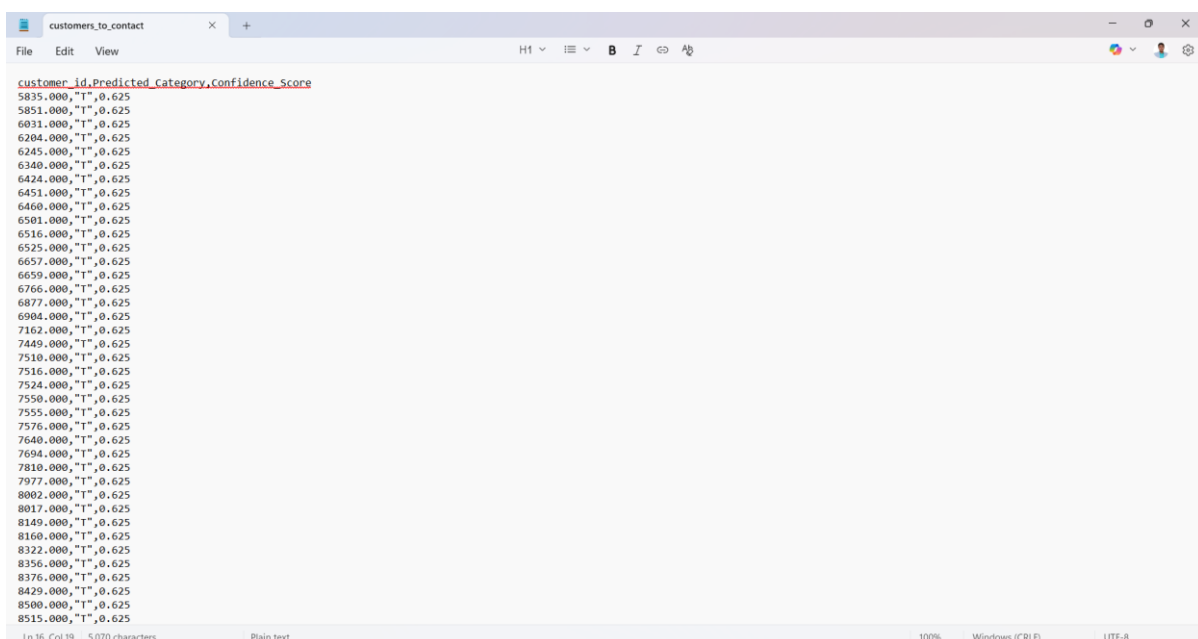
Q 8. How many customers are predicted to respond positively (predicted = T)?

Ans. There are **254** customers are predicted to respond positively (predicted = T).



customer_id	Predicted_Category	Confidence_Score
215	T	0.625
216	T	0.625
217	T	0.625
218	T	0.625
219	T	0.625
220	T	0.625
221	T	0.625
222	T	0.625
223	T	0.625
224	T	0.625
225	T	0.625
226	T	0.625
227	T	0.625
228	T	0.625
229	T	0.625
230	T	0.625
231	T	0.625
232	T	0.625
233	T	0.625
234	T	0.625
235	T	0.625
236	T	0.625
237	T	0.625
238	T	0.625
239	T	0.625
240	T	0.625
241	T	0.625
242	T	0.625
243	T	0.625
244	T	0.625
245	T	0.625
246	T	0.625
247	T	0.625
248	T	0.625
249	T	0.625
250	T	0.625
251	T	0.625
252	T	0.625
253	T	0.625
254	T	0.625

After exporting -



```
customer_id,Predicted_Category,Confidence_Score
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
```

