**Introduction**

The data set includes the records of customer purchasing product X in last 12 month with some other information. Siemens will launch a new product Y similar as the product X next month. The sales team would like to use the analysis results of historic data to identify the target customers for the new product Y. The fields in the data set are interpreted below:

**Index** – customer ID

**Purchase frequency** – how many times the customer purchased product X in last 12 month. 0 means not purchase

**Industry** – which industry this customer belongs to. There are 7 industries represented by number 1 – 7.

**Distributor customer** – if this customer is the distributor’s customer. ‘Yes’ means this customer is the distributor customer, ‘No’ means it is Siemens direct customer.

**Year being customer** – how many years it has being Siemens customer

**Other products** – if this customer purchased other Siemens product lines in the last 12 month

**Satisfaction** – how satisfied this customer with Siemens in last fiscal year. 1 represents very dissatisfied, 3 represents neutral, 5 represents very satisfied

**Awareness of competitor –** the index of the customer awareness of competitor A’s product, score from 1 to 5, higher score means higher level of competitor A’s product awareness. This index is estimated by field sales

**Competitor shares –** the key competitor A’s market share in the area this customer based in. The value is between 1 to 100, estimated by field sales. Higher score means higher market share in this area.

**Question:**

What kinds of customer Siemens should target for product Y.

**Tips:**

* This is an open question without a certain answer. Any reasonable solutions and other interesting findings are welcome.
* This task addresses on the reasonable methods to use for specific propose and data type, and the interpretation of results.
* Do not need the complicated methods, simple is good.
* Please do not waste the time on presentations such as ppt slides. Model results plus key words for narrative is ok.
* Recommend to finish in 1 hour.
* Any other questions, please ask.