**Default Risk and Prediction**

**Purpose:**

The goal of this project is to predict not just which documents are the most likely to default, but also which customer is the most at risk of defaulting. This is because if one of a customer’s orders is likely to default, they all are.

**Tables:**

Included in the data are three tables.

* accounts\_receivable.csv – This is the document level data that includes pertinent order details.
* customer\_data.csv – This is the customer level data that covers customer descriptions and age.
* customer\_sales\_data.csv – This is the monthly historical sales data at the customer level.

**Important Fields**

* DEFAULTED – This is the target field. A 1 means that the order defaulted in the past, a 0 means that it was paid in full.
* SOLD\_TO – This is a join key connecting accounts\_receivable.csv to customer\_data.csv. It does not represent a customer itself, but a location or office that made the order. Larger customers may have many offices or locations.
* GROUP\_KEY – This is the join key connecting customer\_sales\_data.csv to customer\_data.csv. It is the customer level to be used.
* FISCPER – This is the fiscal period, year and month, for the data. It is formatted as YYYY0MM, with a 0 separating the year and month. It is in both the accounts\_receivable.csv and customer\_sales\_data.csv.

**Remaining Fields by Table**

* accounts\_receivable.csv
  + FISCPER – see above
  + SOLD\_TO – see above
  + AC\_DOC\_NO – This is an index to identify individual orders.
  + ITEM\_NUM – This is the number of line items on the order.
  + CREATED\_ON – This is the date that the order was processed.
  + PAYMENT\_METHOD – This is the payment method used on the order.
  + PAYMENT\_TERMS – This is a coded column for our payment agreement.
  + PAYMENT\_BLOCK – This is another coded column referencing our internal processing.
  + CREDIT\_LC – This is the amount that the order was for.
  + DAYS\_ALLOWED – This is how many days they customer was given to pay off the order.
  + LOB – This is our internal line of business (service, construction, electrical, etc).
  + DEFAULTED – see above
* customer\_data.csv
  + SOLD\_TO – see above
  + GROUP\_KEY – see above
  + CUST\_CLASS – This is a coded column used for our internal customer organization.
  + INDUSTRY – This is the primary industry that the customer belongs to.
  + CREDITGRP – This is a coded column we use for organizing customers by credit.
  + MARKET – Similar to INDUSTRY but more granular.
  + SUB\_MARKET – Even more granular than MARKET.
  + REC\_CR\_LM – This is the recommended credit limit we give the customer based on our internal processes.
  + RISK\_CAT – This is our risk assessment of the customer
  + REP\_GROUP – This is a coded field we use for assigning customers to representatives.
  + CUST\_GROUP – This is another, different coded field used for organizing customers internally.
  + CUST\_CREATEDON – This is the date the customer was entered into our system.
  + NAICS – This is a customer reported industry identification.
  + AVG\_D2P – This is the historical, average days to pay for the customer.
* customer\_sales\_data.csv
  + GROUP\_KEY – see above
  + FISCPER – See above, however, it has been offset forward by one period. The effect is that the document is compared to the customer’s sales from the previous month.
  + sumTOTAL\_CUST\_SALES – This is the customer’s total sales by dollar volume.
  + avgTOTAL\_CUST\_SALES – This is the average value of each purchase made by the customer.