# DAT 650 Use Case Descriptions Document

# Credit/Finance Risk

In response to the financial crisis of 2008–2009, the credit branch has been asked to reassess the method used to determine if an application presents a bad credit risk. The example file, Credit\_Data, has data on 1,000 past credit applicants, described by 31 variables. The goal is to determine the likelihood of default for a new credit application based on a subset of the 31 variables. The business user will use this model to help determine the risk associated with extending the credit applicant credit based on the data provided on the application. The impact of the dollar loss to the company on the default of a customer loan is 150% of the remaining balance.

The current data environment is an excel spreadsheet that contains information about the active credit applicants assigned to a business user. The spreadsheet allows the business user adjust the values of certain variables like loan amount to generate the scenario based predictive likelihood that the customer will default. The model will run nightly for all loan applicants and current loan holders with reports generated for high risk loan applicants/holders.

The data is stored in an Oracle database in transactional form. The IT department has built a data warehouse that is updated each evening with the current day data. The Credit team uses this Datawarehouse data as the source to their reports and have the ability to have an adhoc extract to select data into excel for their unique research needs. The Credit Team extracts are limited in rows as well as fields that have been corporately pre approved for extraction.

GE has compiled a file for the purpose of this pilot project using the extract tool. The dataset provided,Credit\_Data.csv, includes information on credit applicants which the GE Credit team believes to be relavent to analyzing this problem.

The Credit team would like to determine if this data can be used to identify Credit default of loan applicants. It is important to be able to understand loan default drivers for meta data like salary, interest rates and other pertinent groupings which comes from the analysis.

The pilot will need to only show basis for this data to be able to describe and generally identify credit applicants that may default. The management team expects to make a GO or NO GO business decision based on the pilot recommendation. If there is a GO, then GE will allocate new project dollars to arrange for GE resources to develop a full enterprise deployed predictivie analytic model. Note that the results of this pilot will be used as a basis for that next project.