Guido,

Thank you for providing the Python tutorial for our use as an introduction to the new software program. This report is an assessment of the tutorial per your request. We also offer some recommendations based on our findings from the Exploratory Data Analysis (EDA).

Installing Anaconda and Python was straightforward and accomplished with ease. No problems occurred and the instructions were complete and correct. Even accessing and using the DOS/command prompt was simple with the clear instructions provided.

Using the book, Python for Data Science for Dummies, was a good way to introduce our team to the basic commands and workings of python. The book I used had several instances of missing parenthesis in the print commands but no other mistakes were noted. That error provided a learning opportunity for me as I am more attentive to command syntax.

The titanic-EDA was extremely helpful in providing a basis for data analysis. We went through the EDA using the titanic dataset and created our own notebook file based on the code in the notebook provided. This allowed us to learn the concepts and code necessary for our initial analysis of our dataset. We explored further by using internet searches to expand on the concepts illustrated in the titanic-EDA.

Our insights found while performing EDA on our credit dataset led to the following recommendations:

* There are 7 entries for Education levels including 4 codes for ‘Other’. We recommend combining others into one code. The EDA reflects that change.
* There are 81 separate Credit\_Limits. We recommend grouping or ‘binning’ them into ranges. A section of code in the EDA shows this view.
* Upon review of defaults in the groups of Credit\_Limit ranges, very few observations are recorded in ranges above $400,000. In the next step we will investigate the effect of removing those observations.
* The feature Age has as similar problem as Credit\_Limit. We also recommend establishing age ranges. Again, code in the EDA show this.
* The ID feature should be removed. Code in the EDA accomplishes this.
* The ‘default payment next month’ should be renamed ‘default’. Code in the EDA does this.
* Our approach in the next step will be to run models on the original dataset. We will also investigate how eliminating entries with Credit\_Limit over $400,000 effects the models. Other features will be examined as well.