Programming Assignment I: using Spark/Pyspark to analyze a bank dataset

1. Use the data dataset posted in the Example-bank dataset UCI folder
   The dataset uses csv with delimiter ';'
   You can use the small {bank.csv} or larger {bank-full.csv} datasets
   The Datasets come from:
   http://archive.ics.uci.edu/ml/machine-learning-databases/00222/
   Description of attributes:
   http://archive.ics.uci.edu/ml/datasets/bank+marketing#

2. The recommended platform {less required settings} is using the Databricks Community edition of Spark online as it is free and most of it is setup with latest version of Spark, Python, SQL, Graph…
   https://community.cloud.databricks.com/login.html

3. If not using the Databricks online you could set your own platform {e.g. using Python and Anaconda to run Spark, Pyspark…. You have to use Spark version > 2.0 in order to have Yarn, Dataframes, ….

4. Produce a set of business hypotheses {8-10} and program them in Python to analyze the data from a variety of angles:
   - Descriptive {how many men/women; Education level; Employed; Married/single/divorced..; Default on loan or not; and combination thereof..
   - Try some prescriptive statistics {predict them failing/not failing to repay their loan} if feasible
   - Try some graphics {bar, pie, line..} if feasible
   - Analyze the marketing results of the previous marketing campaign {any good?}

5. Deliver the programming assignment:
   A. Code in notebook format {Spark – Python}
   B. Screenshots of the code running output {in readable format.. e.g. pdf, or jpg, or png…}
   C. Optional: MP4 run of the code …through screen capture.