**Data Science Take-Home Assessment**

**Timing:**

Please submit your responses according to the directions below within **96 hours**.

**Objective:**

Big Mega, Inc. operates a large chain of stores and has enlisted our help to increase their sales profit. We have received a dataset of historical customer purchases and we are interested in predicting which customers will return to the store again.

**Data source:**

* Shopper’s dataset (shop\_data.csv)

**Tasks:**

1. Predict the probability of a customer returning to the store
   1. Use a random seed or provide indices of train/test/validation splits of customers so your work is reproducible
   2. Use / engineer any features you like and provide commentary on your choices
   3. Evaluate the model using your choice of metric(s) and provide justification for your choice(s)
   4. Optimize the model to your satisfaction and provide commentary on your choice of hyperparameters
2. Presentation / documentation of your process:
   1. Provide all justifications and commentary requested in Task 1 on your thought process – we want to understand WHY you made choices
   2. How does this model help Big Mega accomplish their goal of increasing sales profit?
   3. Do you think this data is appropriate for the objective? What other data, if any, would you request from Big Mega?

**Language requirements:**

* Code must be written in Python, PySpark, or Scala (2.11)
* Code may be written in a Jupyter Notebook (\*.ipynb), Python script (\*.py) or a script that would be used in a spark-submit (Spark 2.4.3)

**Submission and scoring:**

* Put all needed code and responses in a zip file named:
  + firstname\_lastname\_CerebriDStest.zip
* Submission will be graded on:
  + Code quality
  + Model output including feature engineering and performance results
  + Justifications/commentary on feature engineering and hyperparameter choices
  + Responses to 2.b. and 2.c., above
* Bonus points will be awarded for coding in PySpark with more bonus points awarded for coding in Scala