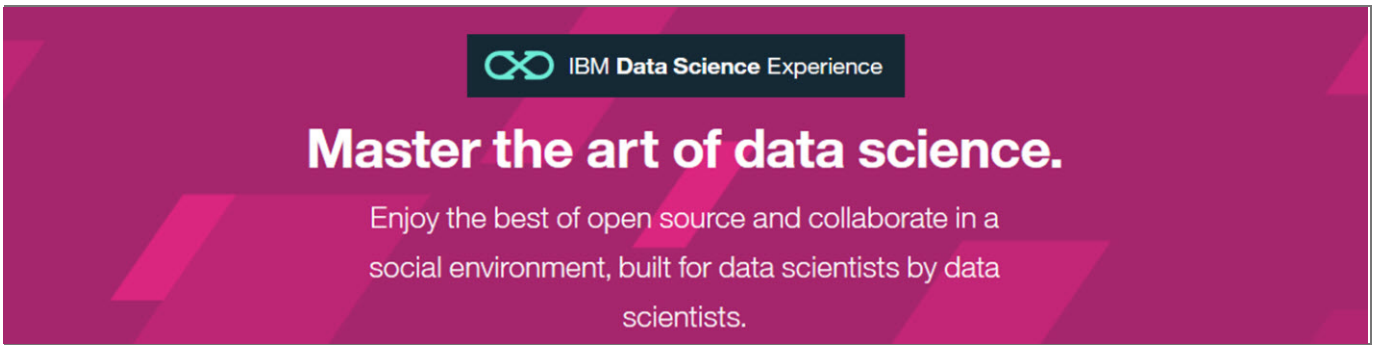


Data Science Experience Local on IBM Integrated Analytics System Backup Hands-On Lab

IBM NA Big Data Science Technical Team

October 2017

Contents

INTRO **IBM DATA SCIENCE EXPERIENCE LOCAL ON IBM INTEGRATED ANALYTICS SYSTEM..... 3**
PART 1 JUPYTER NOTEBOOK THAT PREDICTS OUTDOOR EQUIPMENT PURCHASES 3
PART 2 USING THE MACHINE LEARNING BUILDER 6

Intro IBM Data Science Experience Local on IBM Integrated Analytics System

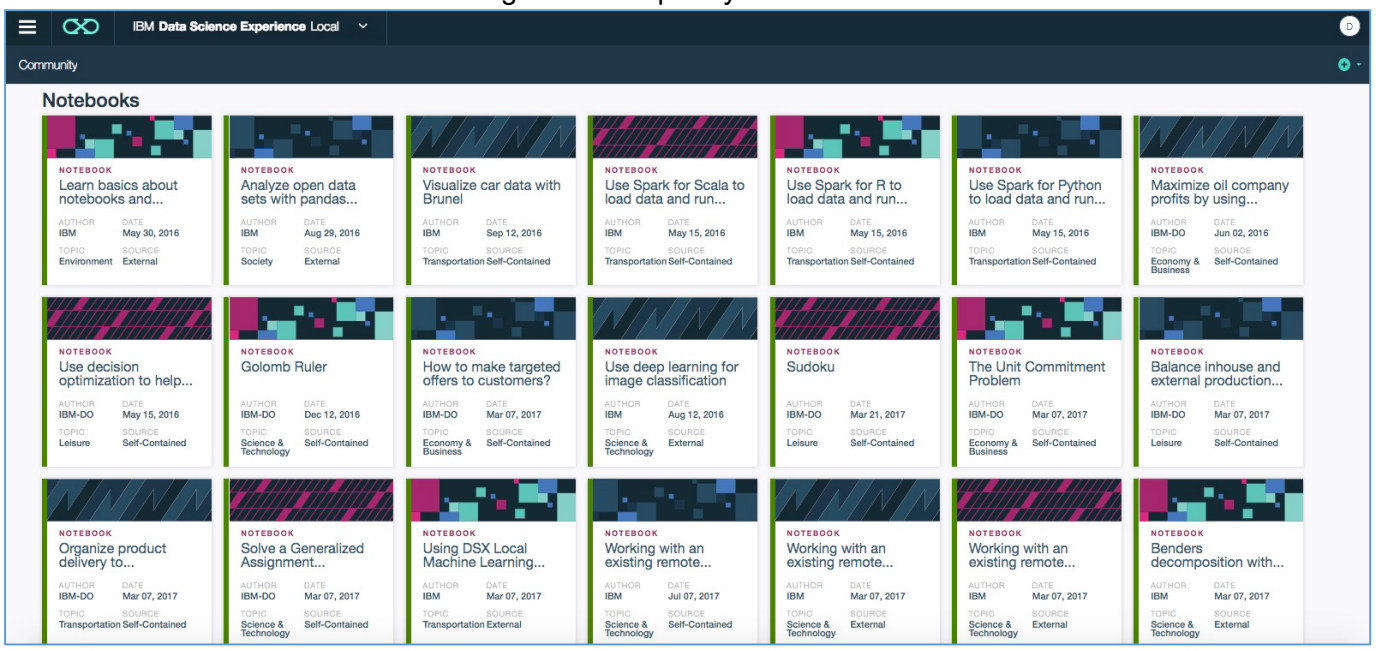
IBM Data Science Experience (DSX) Local is an out-of-the box on premises enterprise solution for data scientists and data engineers. It offers a suite of data science tools that are integrated with proprietary IBM technologies. It also seamlessly integrates with RStudio, Spark, Jupyter and Zeppelin notebook technologies. The intuitive user interface provides a collaborative Projects space for teams and individuals to reduce the time to value.

Part 1 Jupyter Notebook that Predicts Outdoor Equipment Purchases

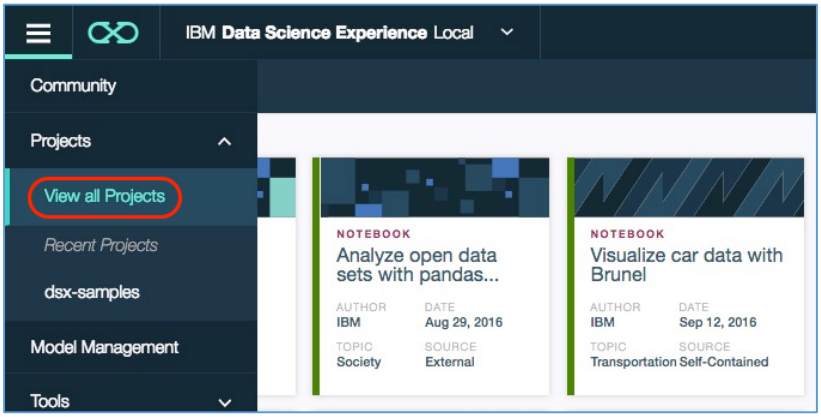
This part of the lab will import a Jupyter notebook that will predict product line purchases based on a consumer's gender, age, marital status and job. This notebook will show you how to load and explore data, create a machine learning model, persist the model, predict locally and visualize, deploy and score the model.

a) Open your browser and navigate to http://ibm.biz/dsxlocal to access the DSX Local sign in screen. Use your 'team##' with a password of 'sailfish##', where ## is your team number.

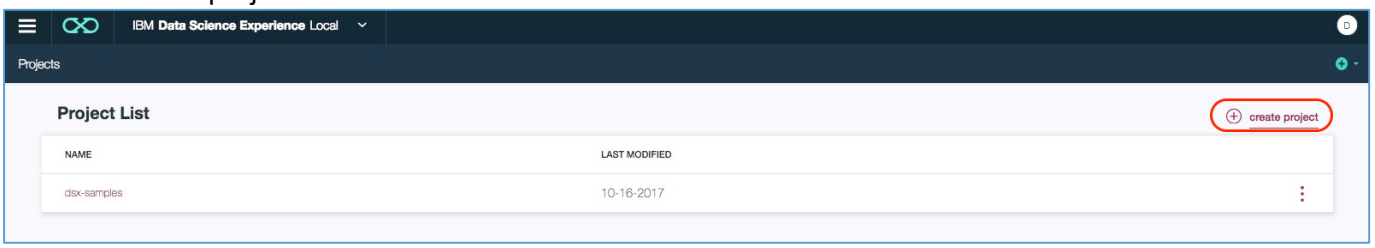
b) After a successful sign in, the following Community screen will appear. Please take a few moments to view the various assets available to get started quickly.



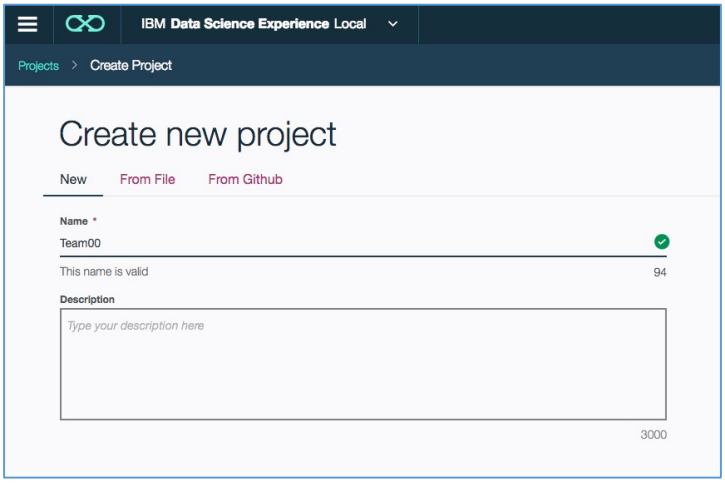
c) Now click on the hamburger icon on the top left, then click on ‘Projects’ and ‘View all Projects’.



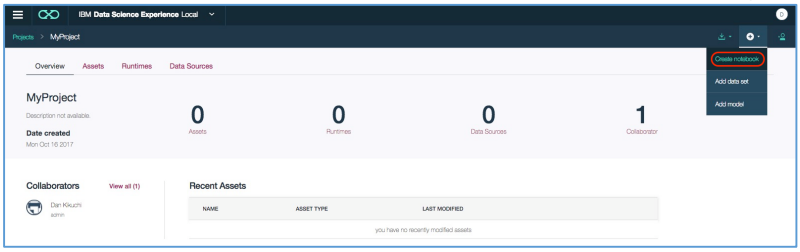
d) Click on ‘create project’.



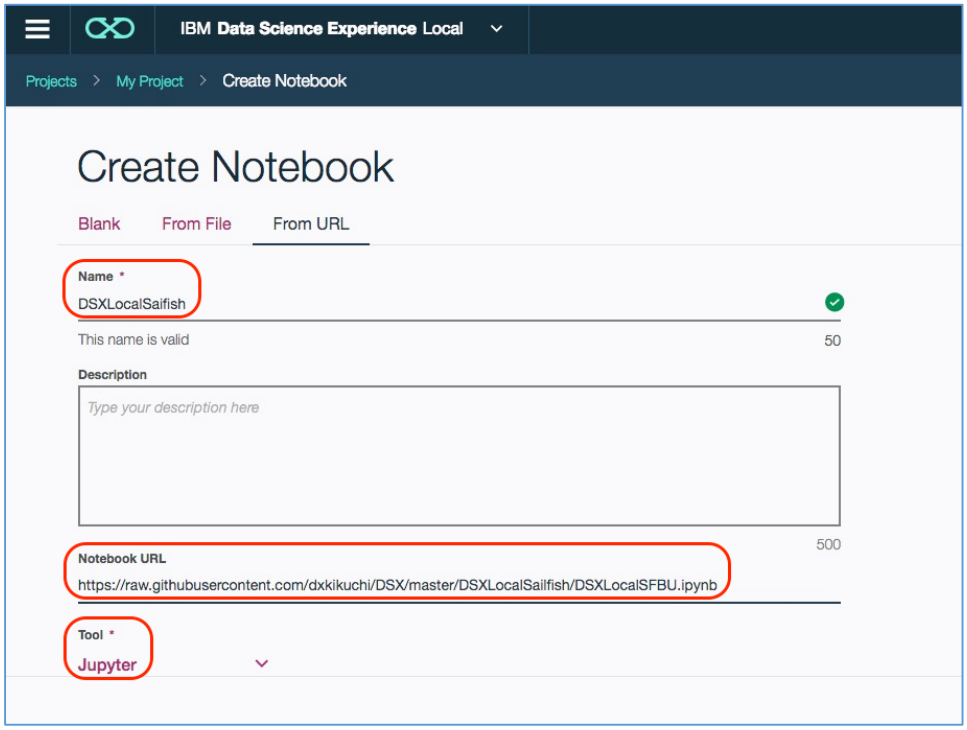
e) Then enter your team name, ‘Team##’ for ‘Name’ and click on ‘Create’.



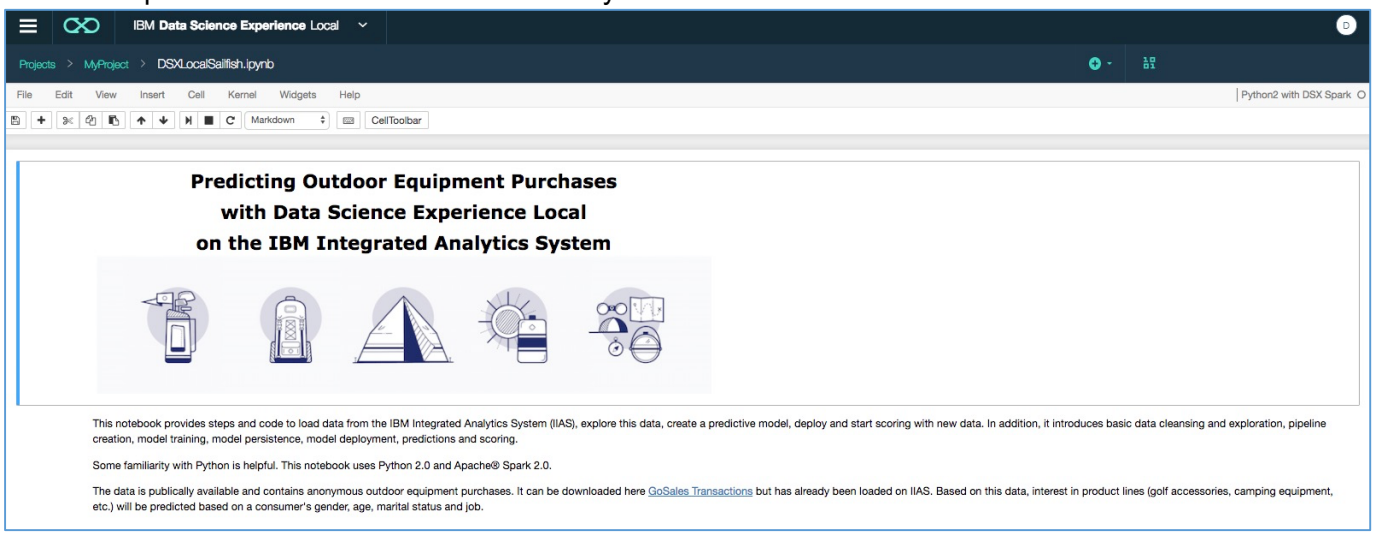
f) Click on ‘Create notebook’ from the pull down in the top right ‘+’ icon.



g) In the ‘Create Notebook’ page, click on ‘From URL’, then enter ‘DSXLocalSailfish’ for the ‘Name’, select ‘Jupyter’ for ‘Tool’, and enter the following for the ‘Notebook URL’.
https://raw.githubusercontent.com/dxkikuchi/DSX/master/DSXLocalSailfish/DSXLocalSFBU.ipynb
Then click on ‘Create’.



h) This will import the notebook from GitHub into your DSX Local environment.

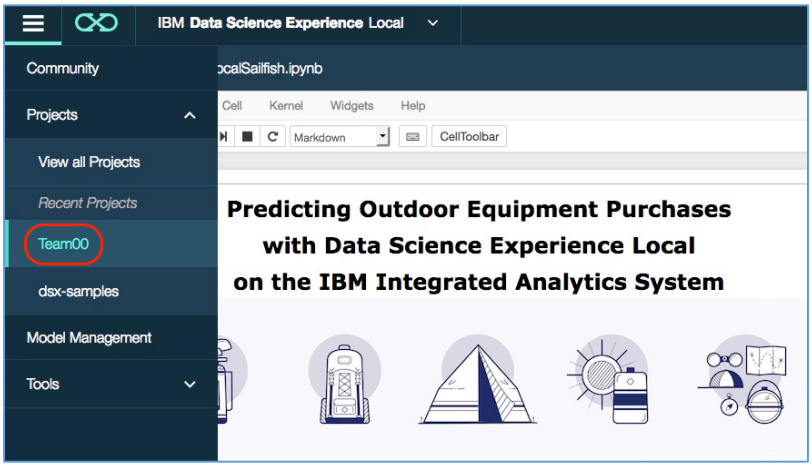


i) Please walk through the notebook by reading and executing the code cells to understand what is being accomplished.

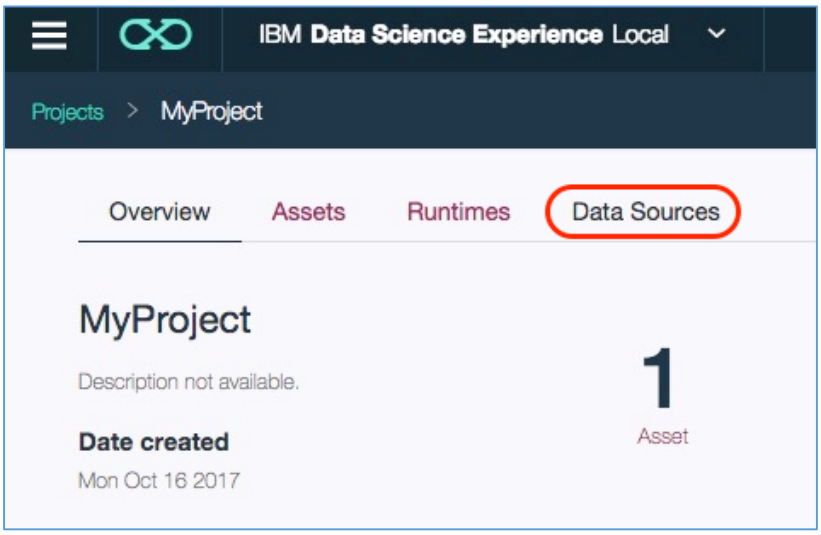
Part 2 Using the Machine Learning Builder

The following will walk through the steps to use the machine learning builder wizard.

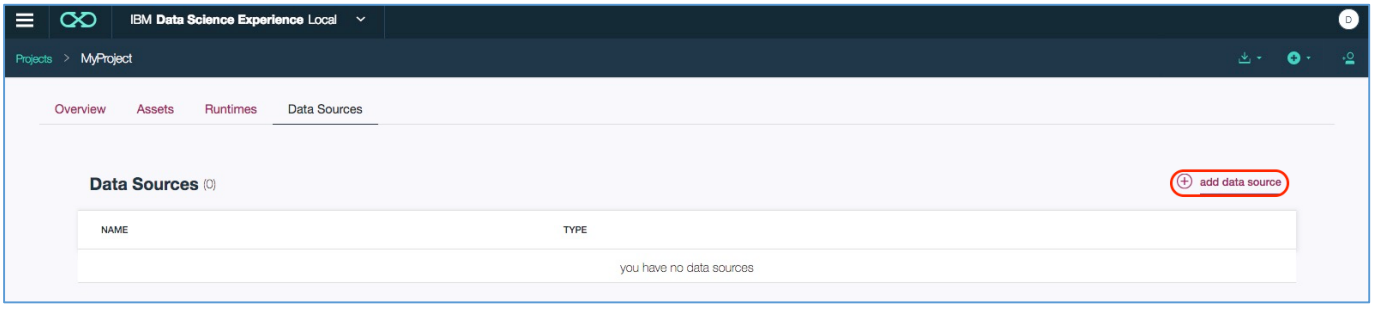
j) Click on the hamburger icon on the top left and select ‘Projects’ then your team project, ‘TeamXX’.



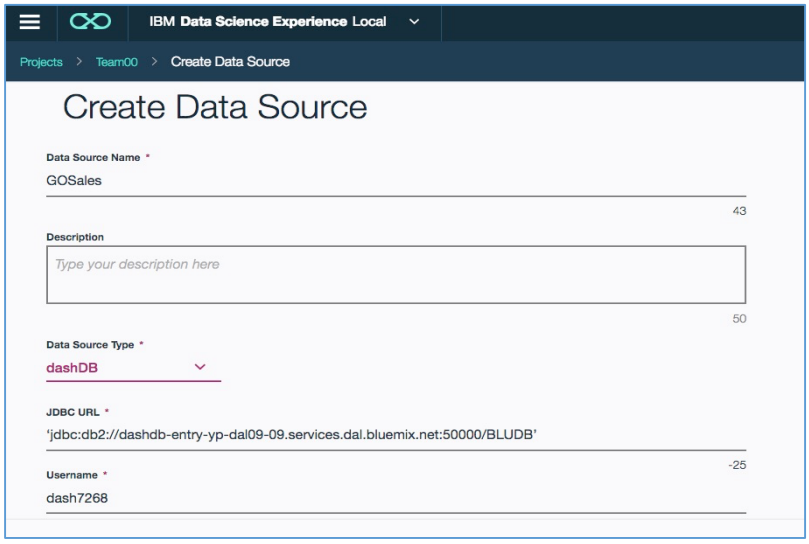
k) First create a data source that will be recognized by the machine learning builder. Click on ‘Data sources’.



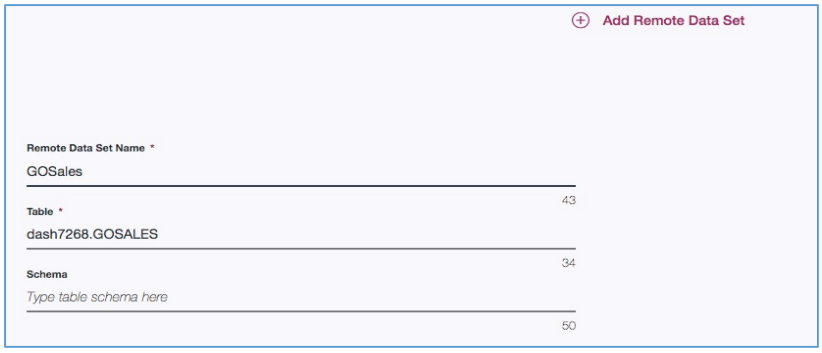
l) Click on ‘add data source’.



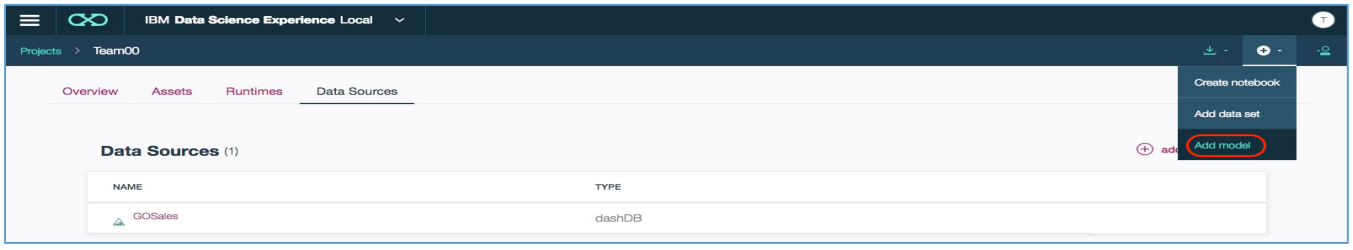
m) On the first part of this screen enter:
‘GOSales’ for the ‘Data Source Name’
‘dashDB’ for the ‘Data Source Type’
‘jdbc:db2://dashdb-entry-yp-dal09-09.services.dal.bluemix.net:50000/BLUDB’ for the ‘JDBC URL’
‘dash7268’ for the ‘Username’
‘mV@$e7q4RPzL’ for the ‘Password’ and
Keep ‘Shared’ UNCHECKED.



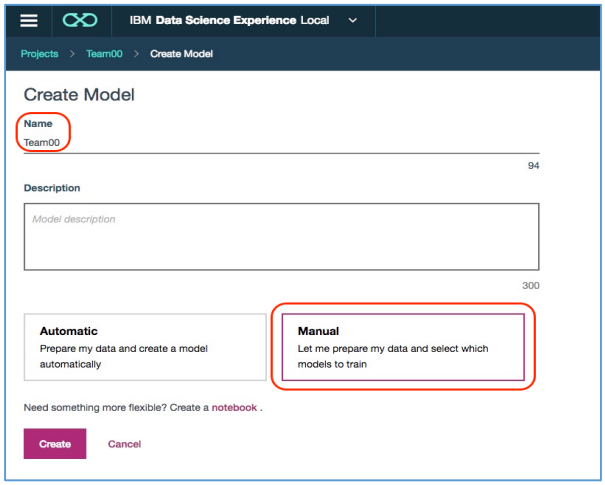
n) Now scroll down, click on ‘Add Remote Data Set’ and enter the following:
‘GOSales’ for the ‘Remote Data Set Name’
‘dash7268.GOSALES’ for the ‘Table’
Leave ‘Schema’ blank.
Now click ‘Create’.



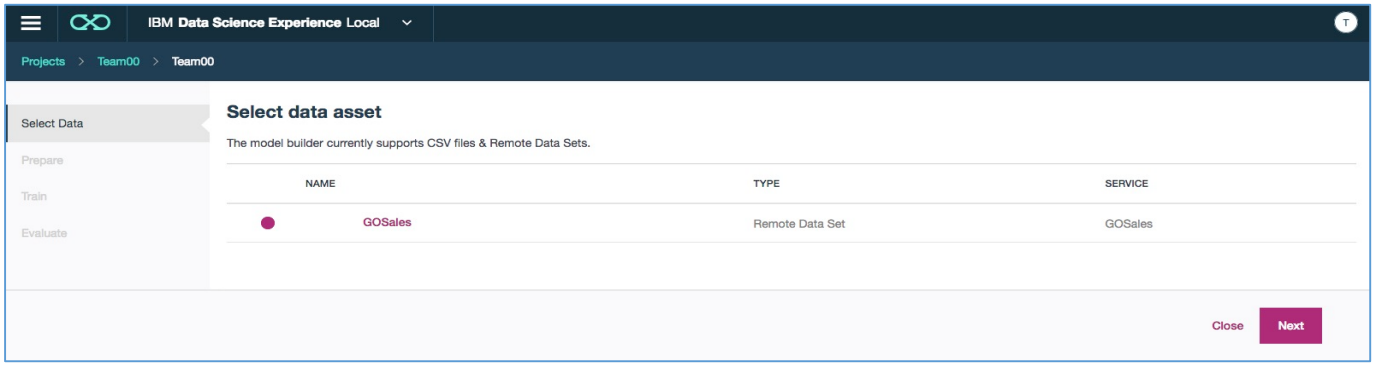
o) The ‘GOSales’ data source should now appear.
Click on ‘Add model’ under ‘+’ pull down on the top right.



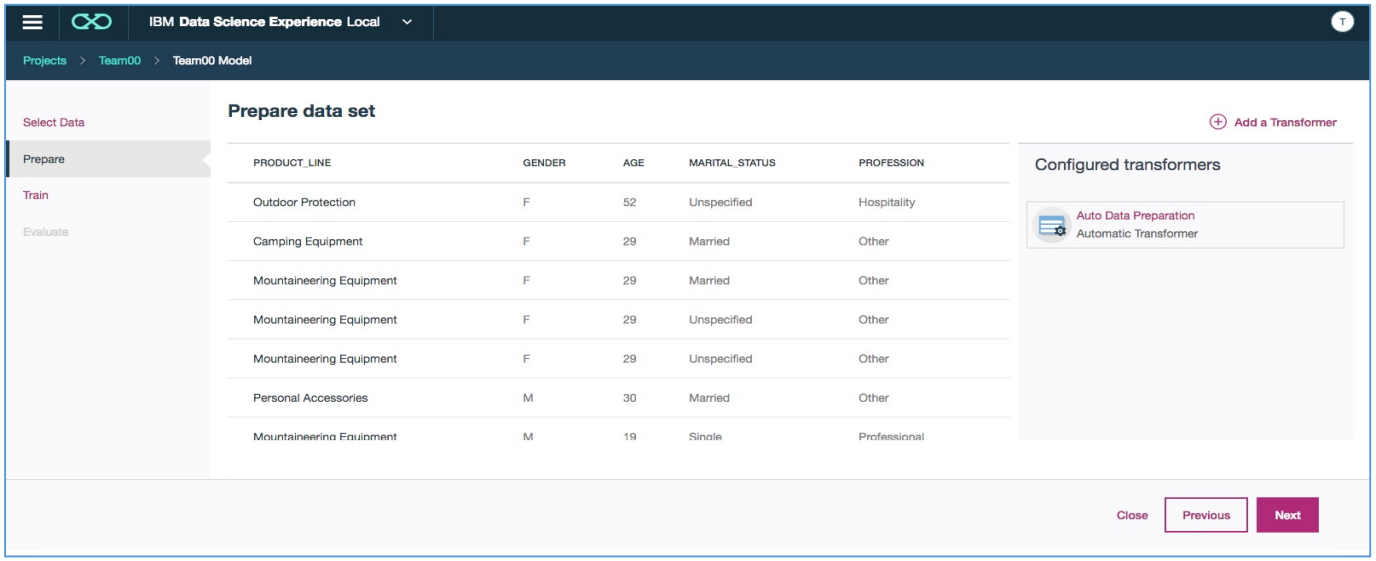
p) On the ‘Create Model’ screen, enter your team name ‘TeamXX’ for the ‘Name’ and select ‘Manual’. Please note that it states ‘Need something more flexible? Create a notebook’. This was done in Part 1 of this lab. Now click on ‘Create’.



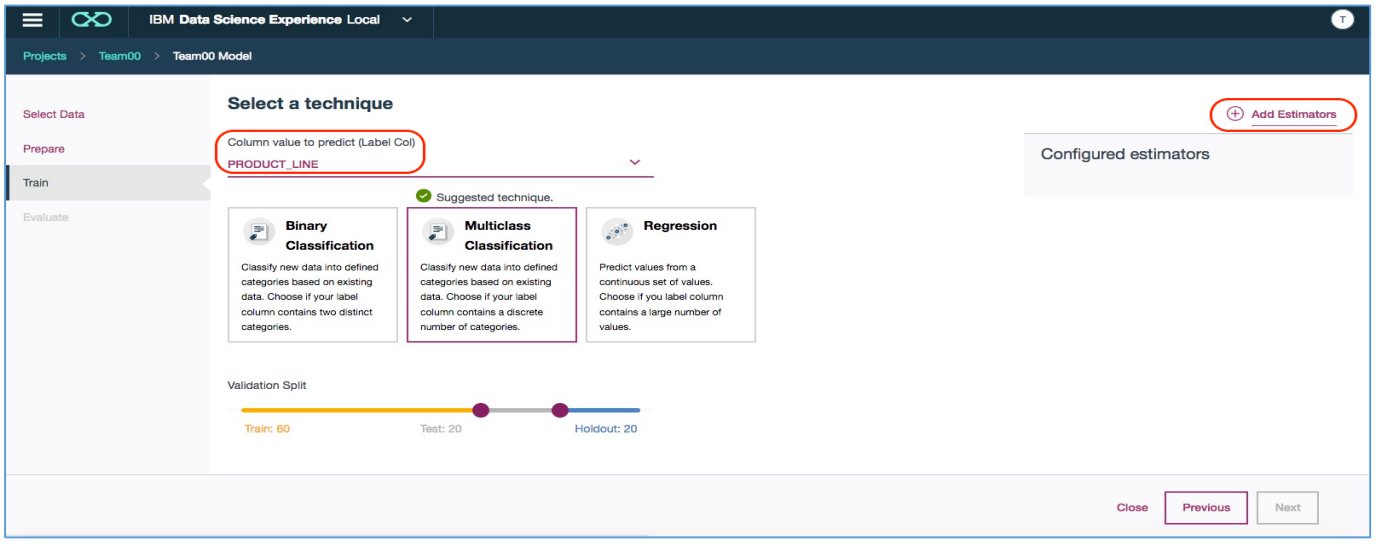
q) Select the ‘GOSales’ data source that we just created and click on ‘Next’.



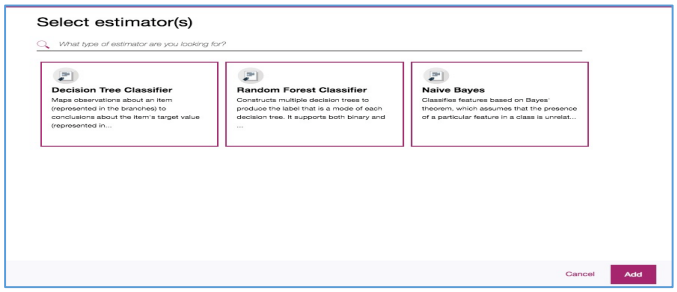
r) Wait until the data is loaded. A small sample of the data will be displayed. Please keep ‘Auto Data Preparation’ and click on ‘Next’.



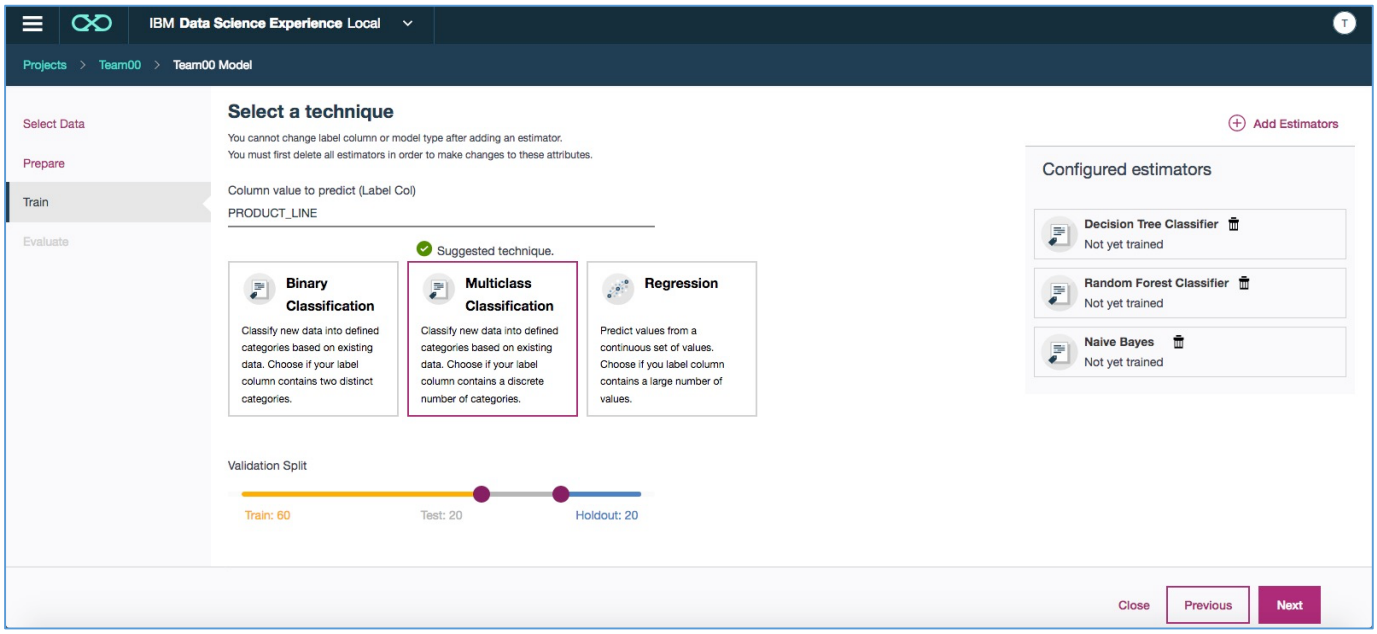
s) Now select ‘PRODUCT_LINE’ from the ‘Column value to predict (Label Col)’. Notice that a ‘Multiclass Classification’ is indicated for the ‘Suggested technique’. Now click on ‘Add Estimators’.



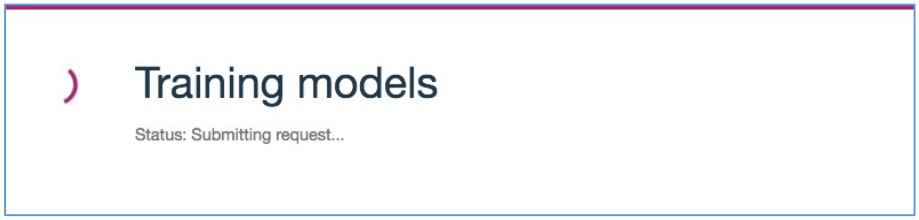
t) Select all three (Decision Tree Classifier, Random Forest Classifier, Naïve Bayes) for comparison. Now click ‘Add’.



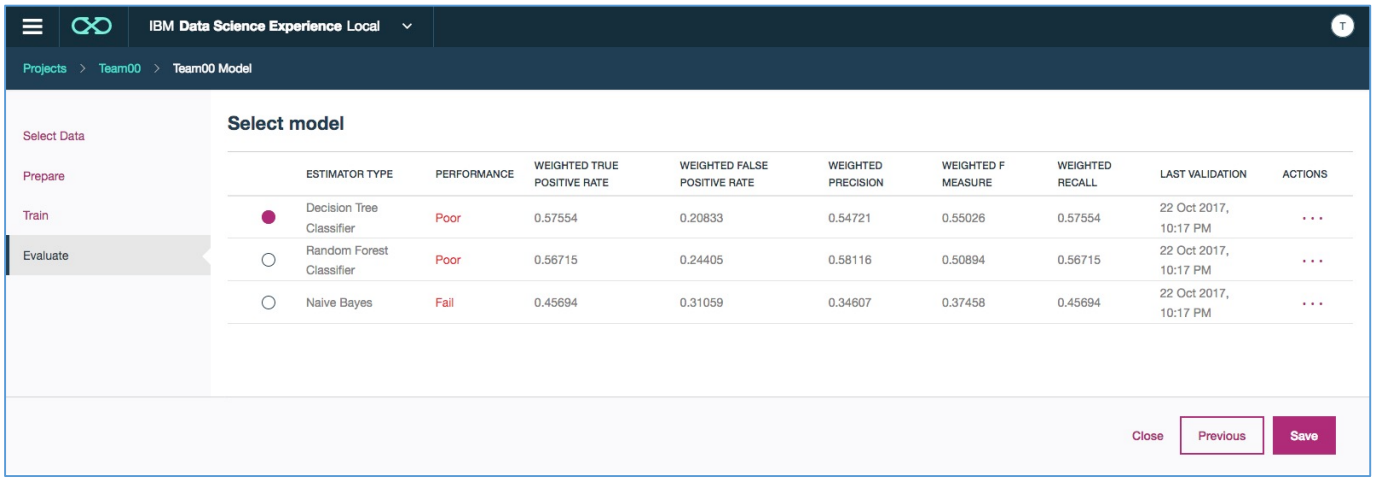
u) Now click on ‘Next’.



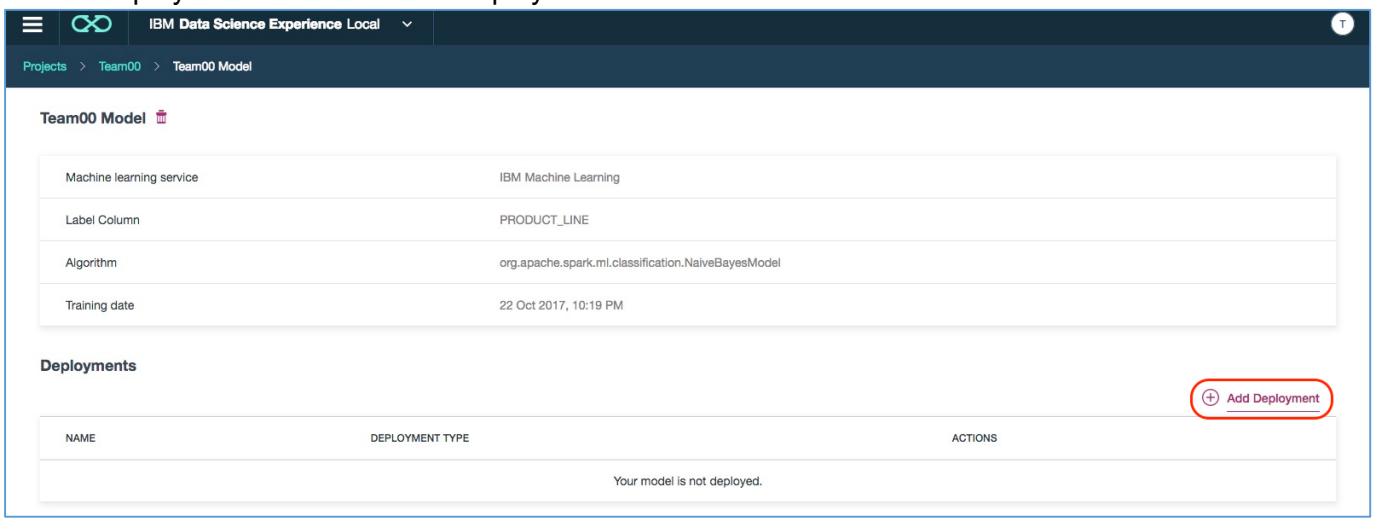
v) The models will now be trained. This may take some time so please be patient.



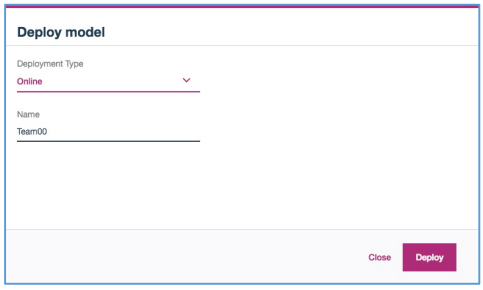
w) The results of the 3 estimators will be listed. In this case, the 'Decision Tree Classifier' provided the best results. Please select the estimator with the best performance and click on 'Save'. Then click on 'Save' again in the pop-up window.
Note: The accuracy of these models is 'Poor' or 'Fair'. Greater accuracy may be achieved by using a notebook as in 'Part 1' and provide more data, regularization, normalization or other techniques. For purposes of this lab to show the process, we will continue with the best of the 3.



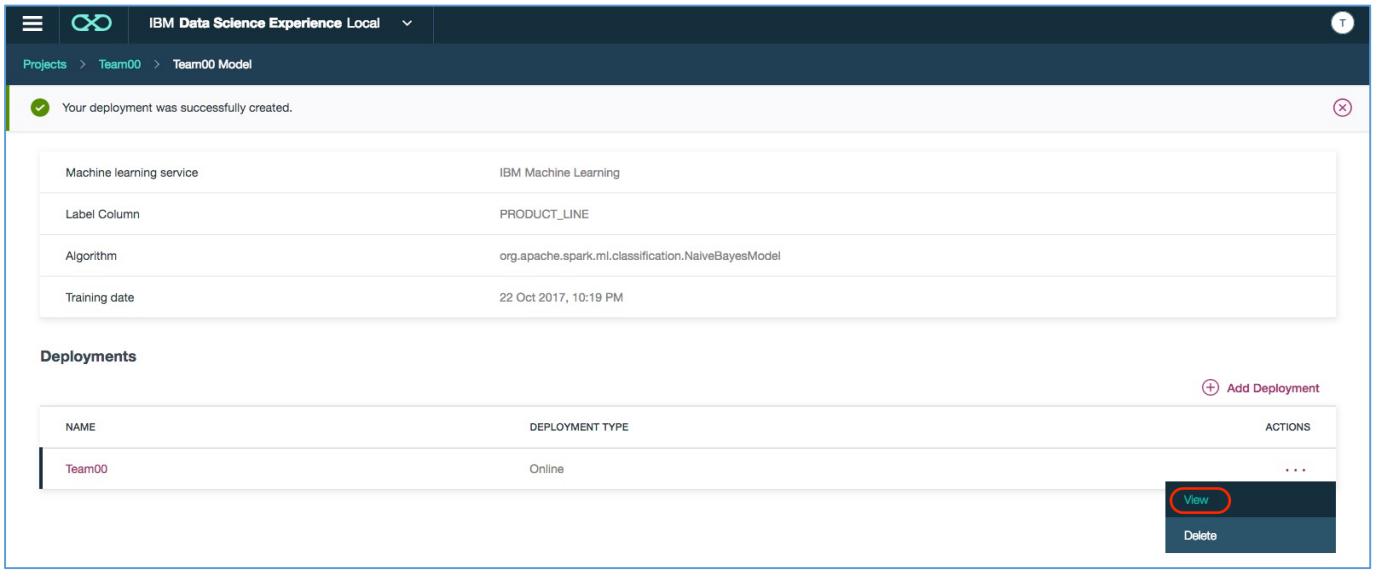
x) Under 'Deployments' click on 'Add Deployment'.



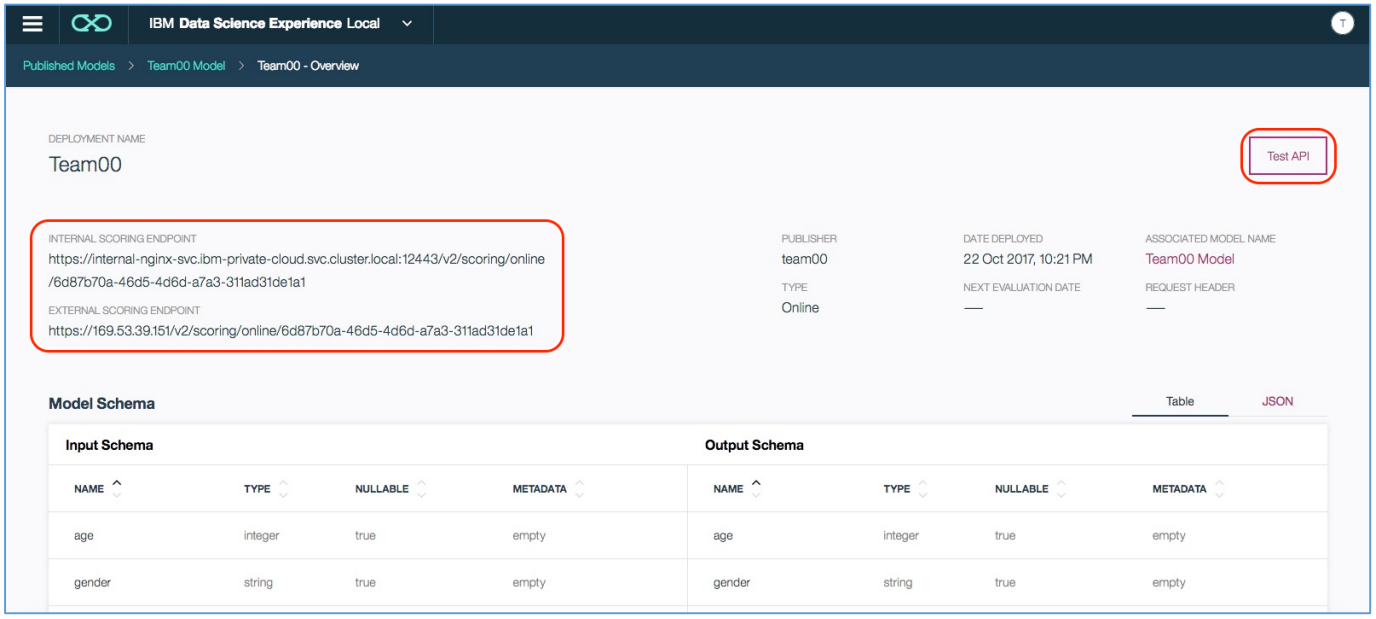
y) In the ‘Deploy model’ screen, select ‘Online’ for ‘Deployment Type’ and your team name ‘TeamXX’ for the ‘Name’. Now click on ‘Deploy’.



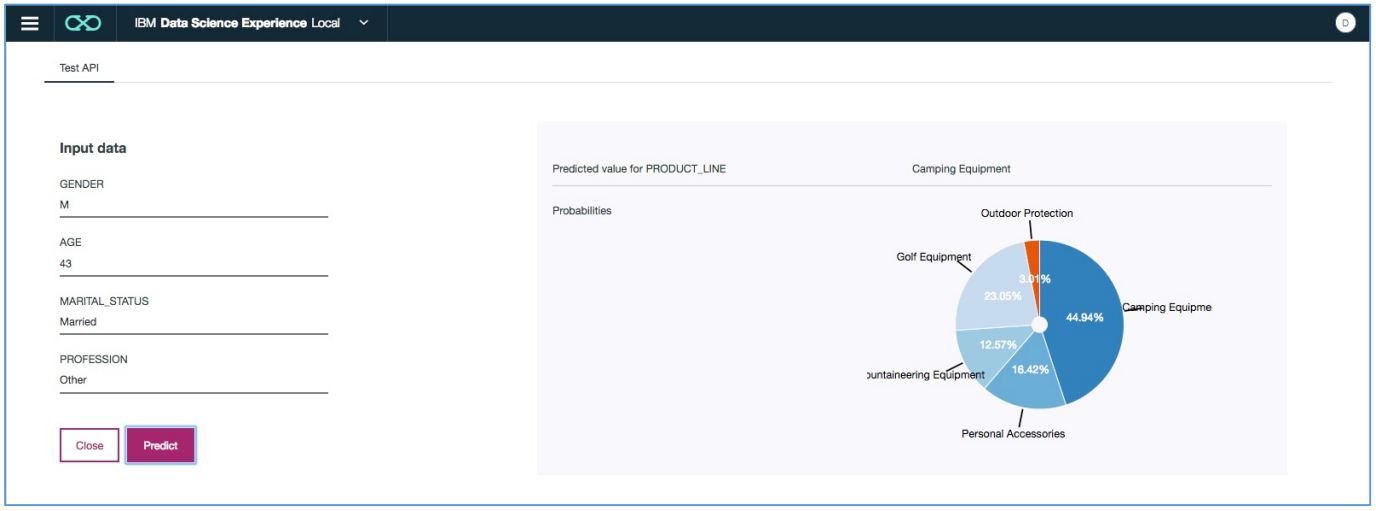
z) Under ‘Deployments’ for your team name, click on ‘View’.



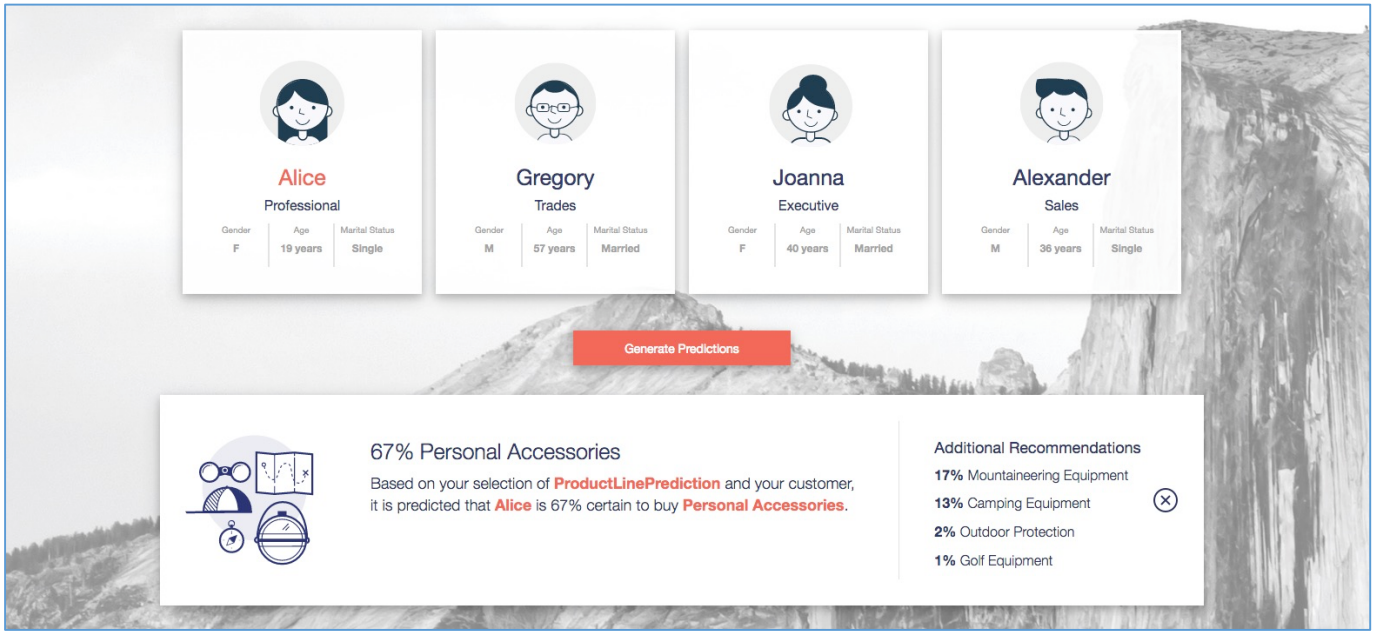
aa) Notice that the scoring endpoints are displayed. Click on ‘TestAPI’.



bb) Use the provided input or specify something different, then click on ‘Predict’. The scoring results will be displayed.



cc) As an aside, the scoring endpoints could be used in a web application to provide real-time scoring.



Congratulations, you have completed this exercise.

Great work and congratulations, you have completed this exercise.

Screen captures in this document may vary slightly from yours.

NOTES

Please Recycle