

Overview

This lab will introduce the Watson Machine Learning capability using the Bike Share dataset. The lab will consist of the following steps:

1. Setting up the environment
2. Adding a data asset to the DSX Labs project
3. Creating a Model to predict the Ride Count
4. Deploying the Model

Step 1: Setting up your environment

To use IBM Watson Machine Learning you must have the following service instances in your Bluemix dashboard:

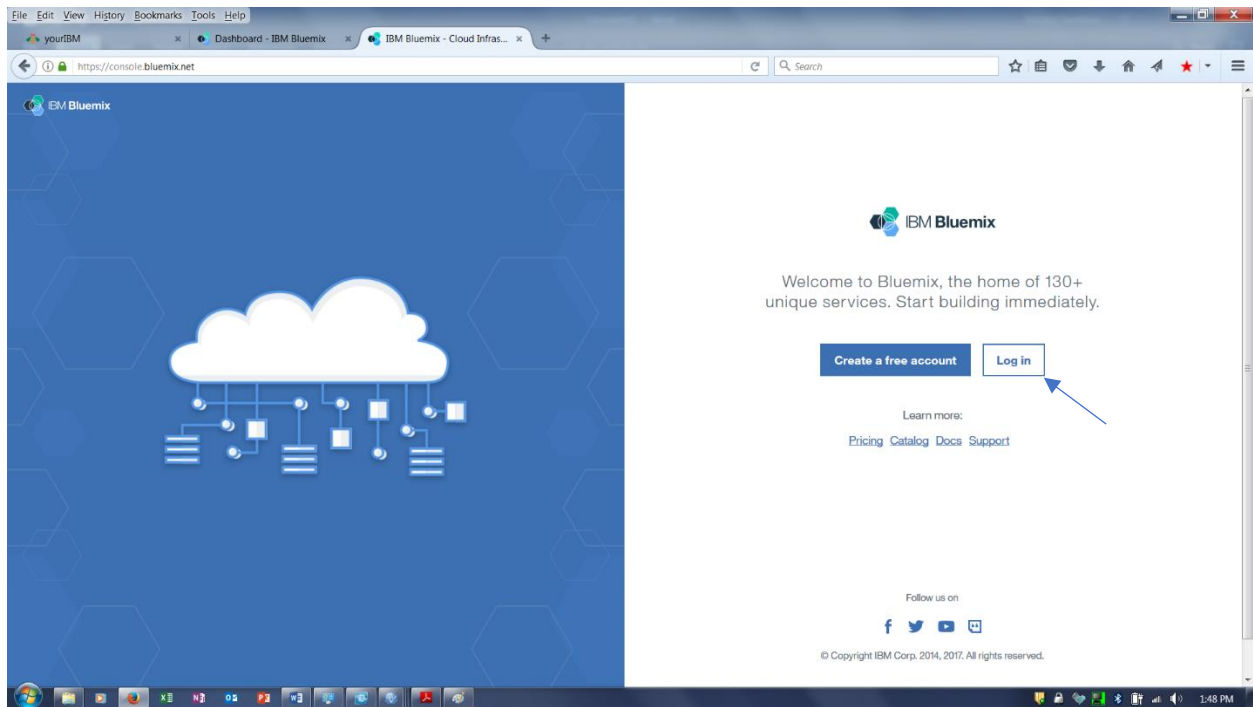
- Watson Machine Learning
- Object Storage
- Apache Spark

The Object Storage and Apache Spark service instances should already exist having been created when your DSX account was provisioned. We now need to provision a Machine Learning Service.

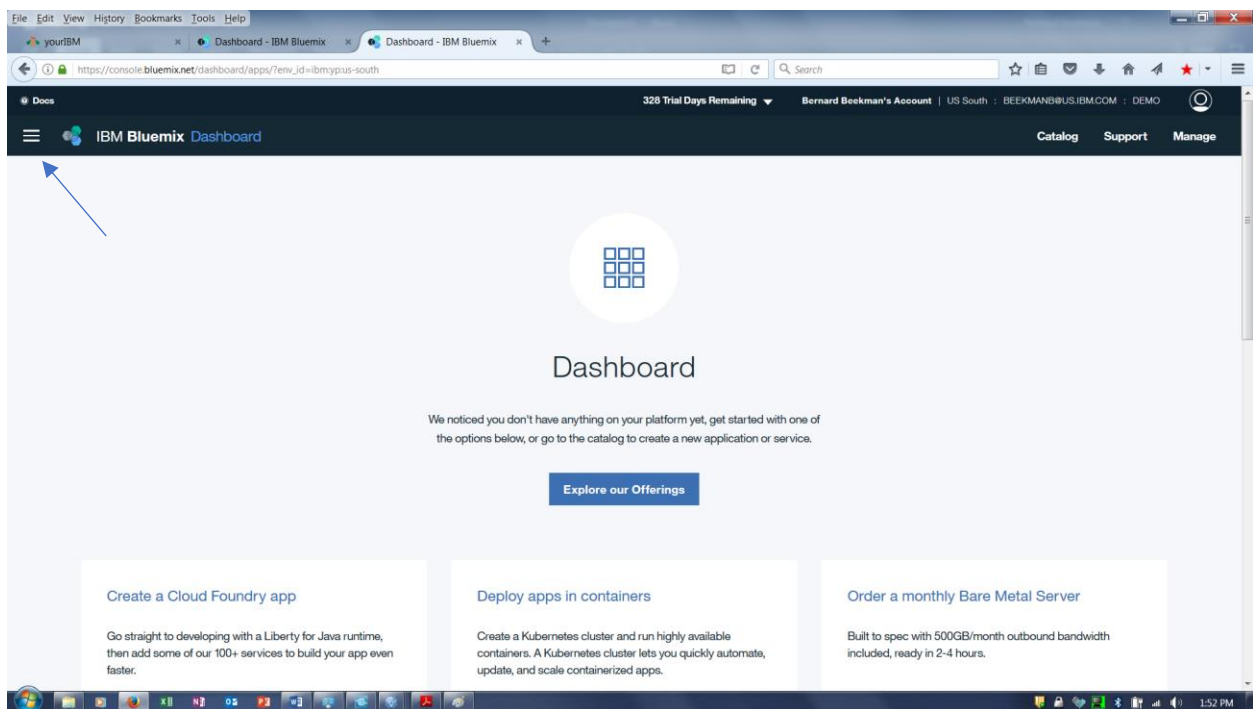
Step 1.1: Creating a Machine Learning Instance

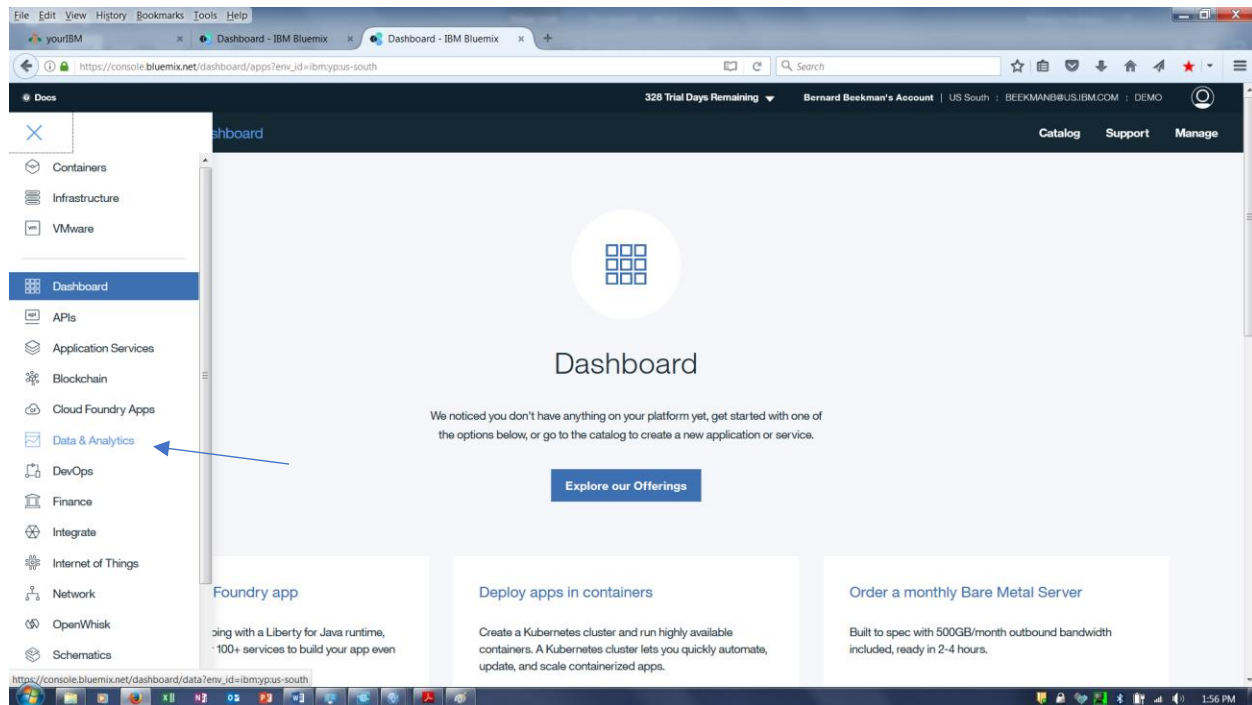
To create a Machine Learning service instance, you must perform the following steps:

1. Log into Bluemix at www.bluemix.net.



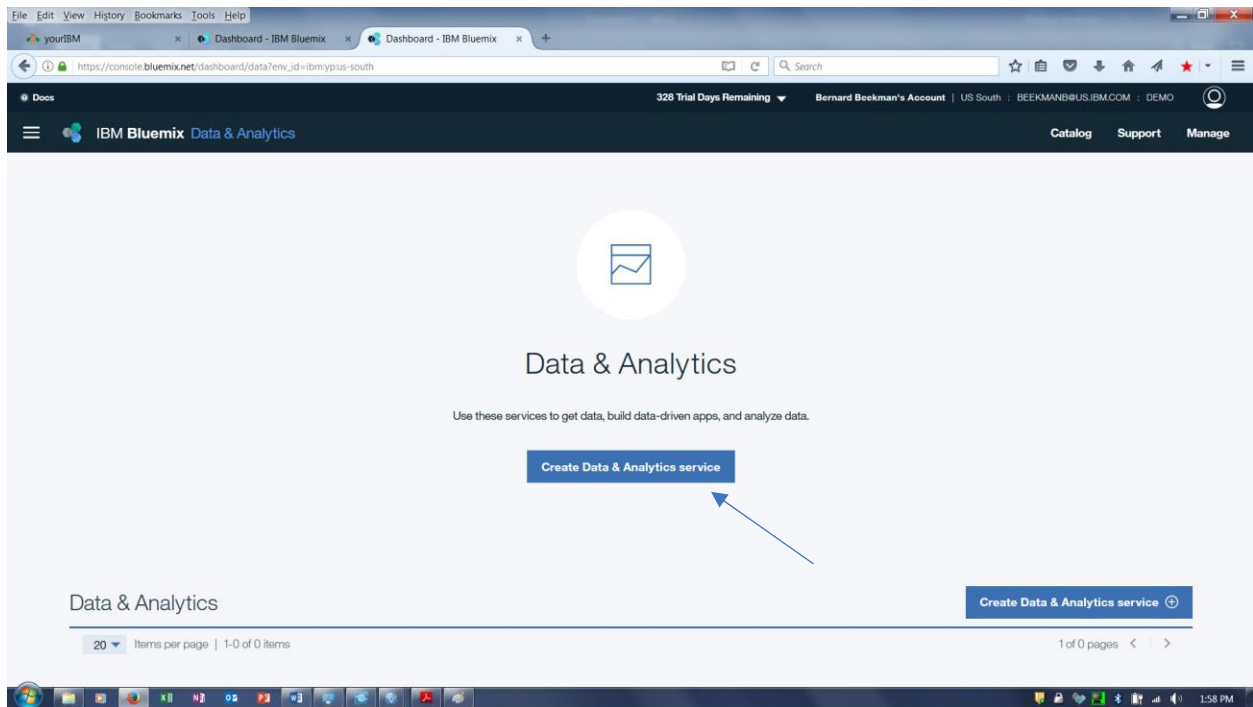
2. Once logged in, click on the hamburger icon, and from the navigation panel, click **Data & Analytics**.



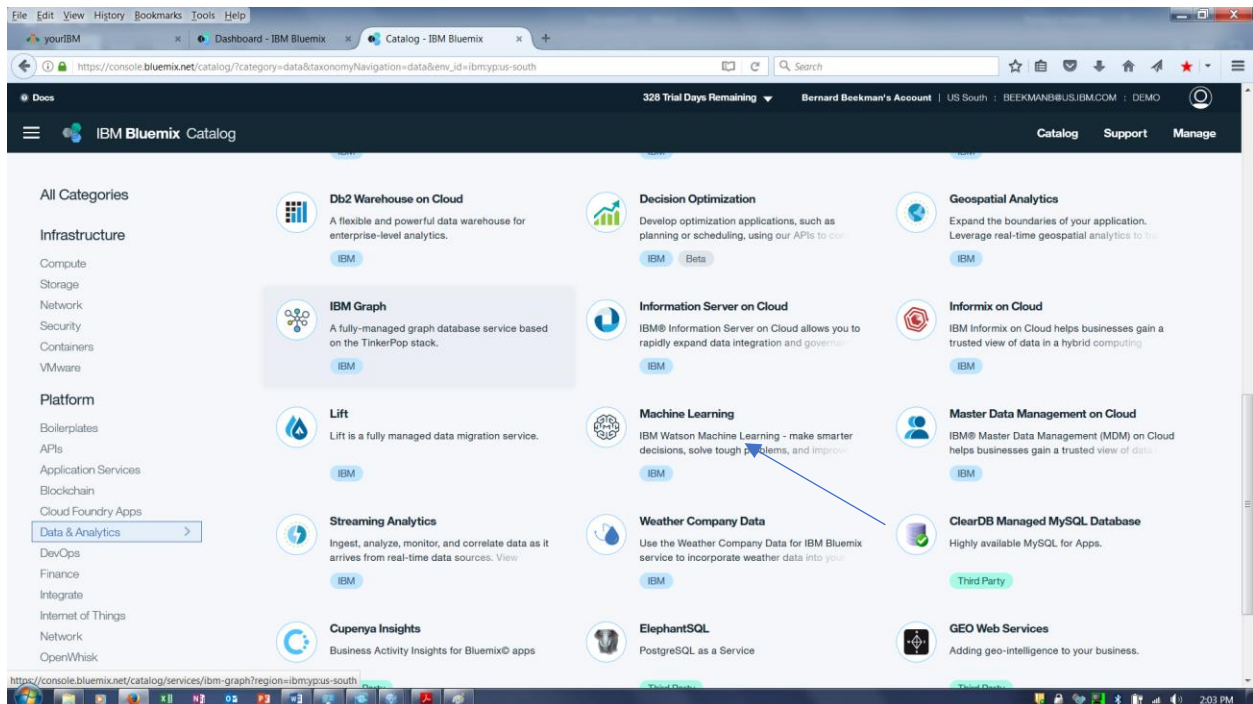


You see a screen centered on data services. You can return here regularly to work with your data and analytics services from one easy-to-use page. Check to see if a Machine Learning service already exists. If not, continue, otherwise go to Step 1.2: Adding existing Bluemix instances to a project in Data Science Experience

3. Click the **Create Data & Analytics Service** button.



4. Scroll down to Machine Learning and click.



5. Configure service.

Enter a descriptive name for your service, choose a space, and select your data plan (find plan comparison and pricing details on this page). Click on **Create**.

The screenshot shows the IBM Bluemix Catalog interface for creating a new 'Machine Learning' service. The page includes a sidebar with 'View all' and 'View Docs' links, and a table of service details (Author: IBM, Published: 08/01/2017, Type: Service, Location: US South, United Kingdom). The main form contains the following fields:

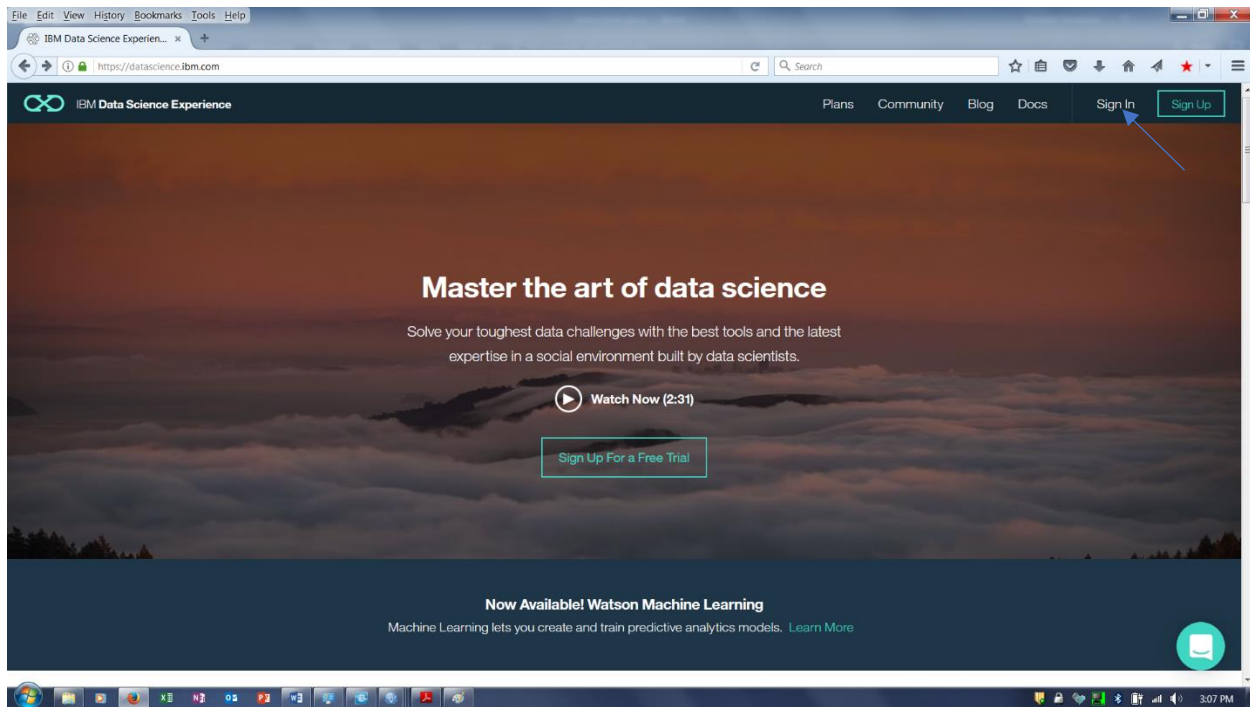
- Service name:** Machine Learning
- Credential name:** Credentials-1
- Select region to deploy in:** US South
- Choose an organization:** BEEKMANB@US.IBM.COM
- Choose a space:** DEMO
- Connect to:** Leave unbound

At the bottom right, there is a blue 'Create' button. Blue arrows in the image point to each of these fields and the 'Create' button.

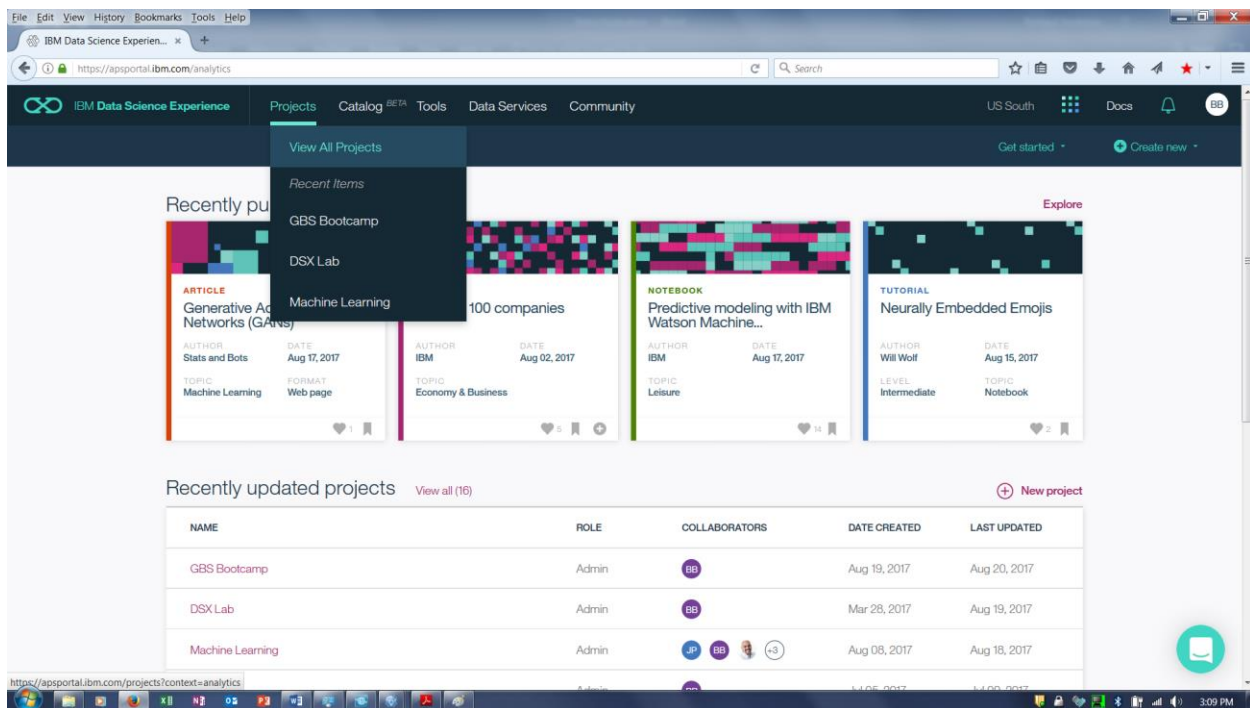
Step 1.2: Adding existing Bluemix instances to a project in Data Science Experience

If you already have instances, but have not linked them to a project in Data Science Experience, you must perform the following steps:

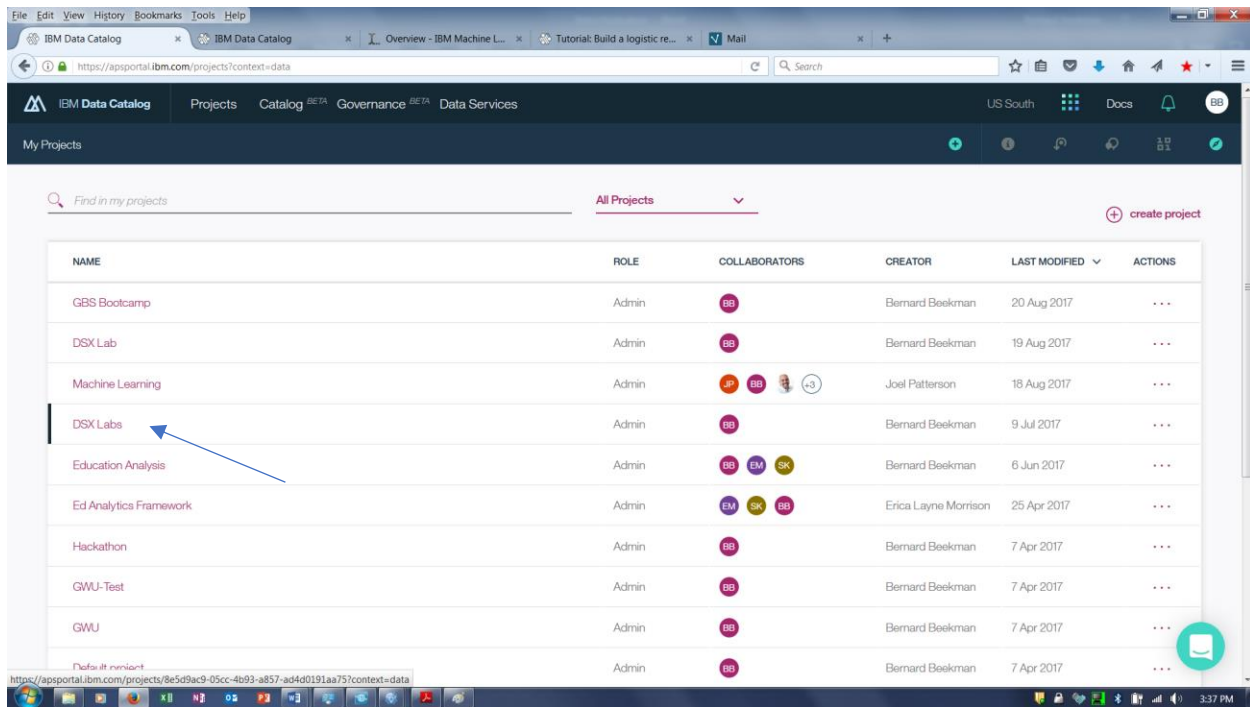
1. Log on to IBM Data Science Experience – <https://datascience.ibm.com>



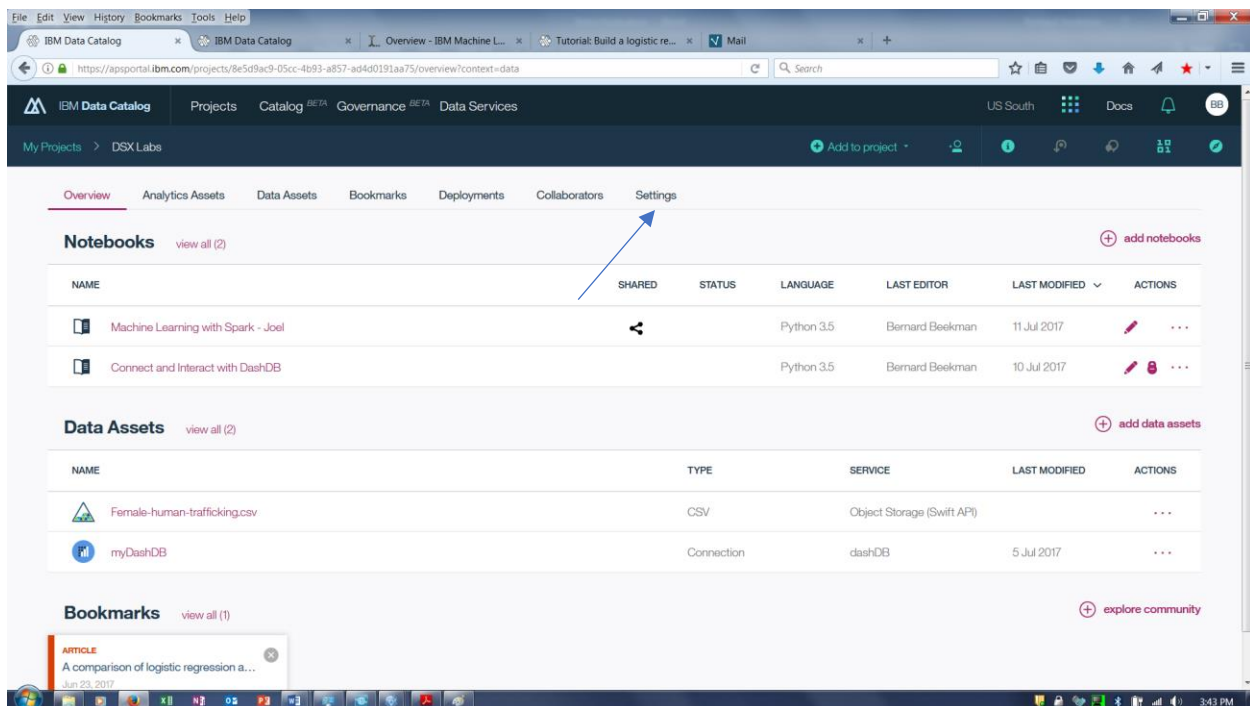
2. Click **Projects** > **View All Projects**.



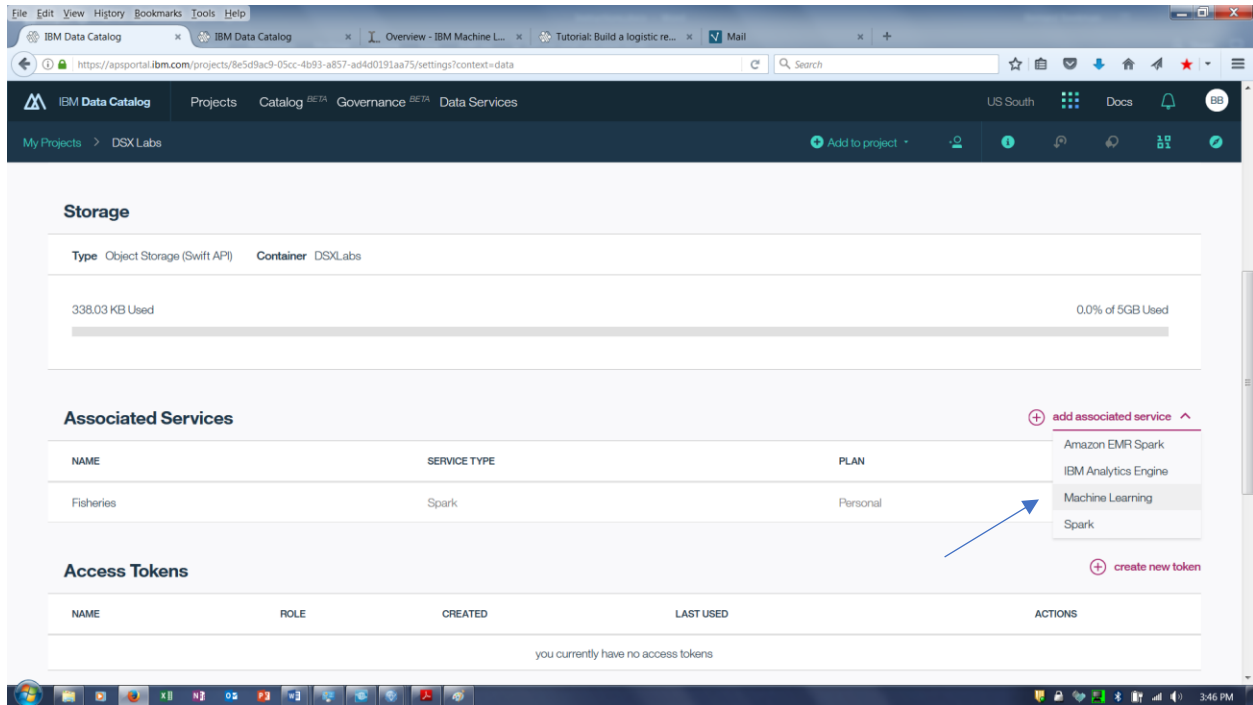
3. Click on the project that you created in the prerequisites – DSX Labs.



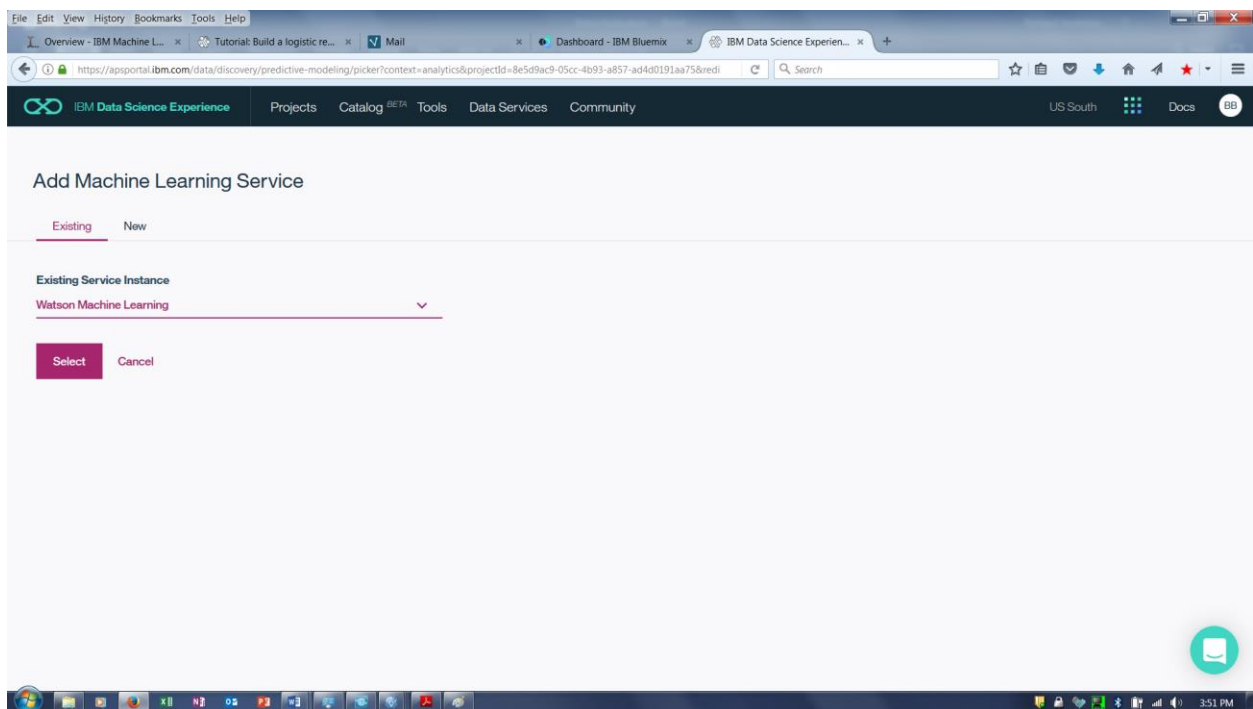
4. Select the **Settings** Tab.



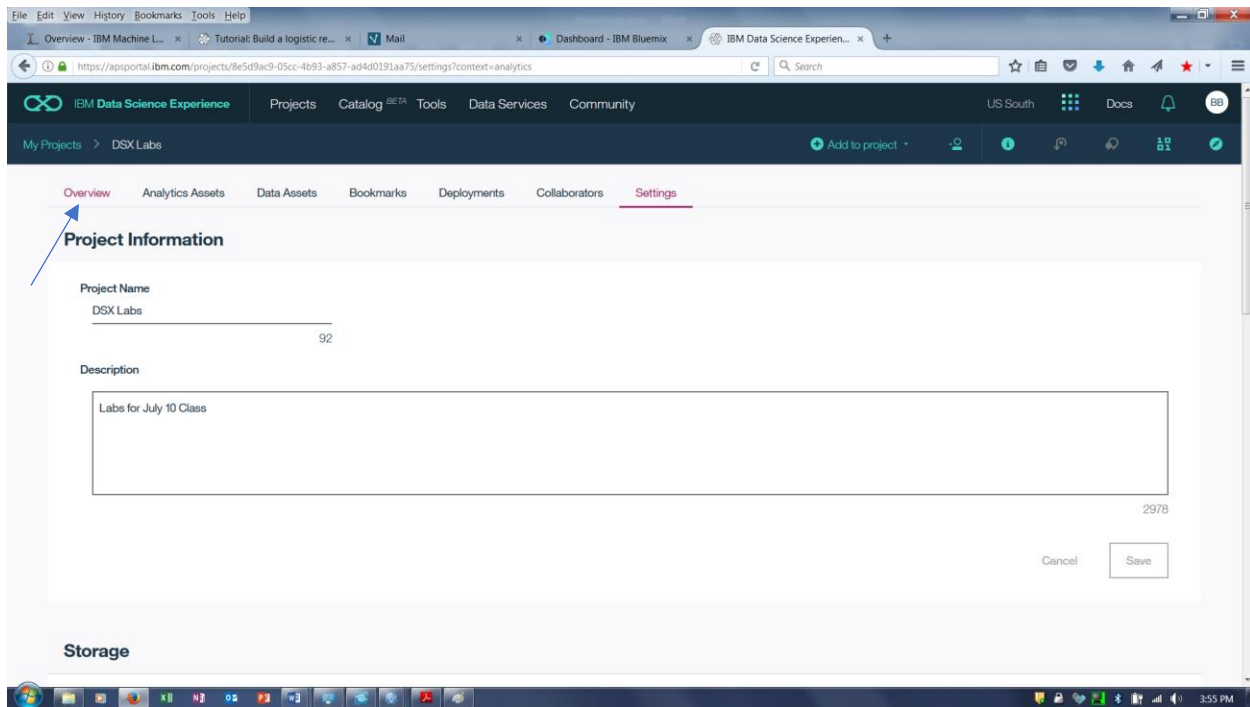
5. Scroll down to Associated Services. To add a service, in the **Associated Services** panel, click **add associated service**, select the Machine Learning service.



6. Select the Machine Learning service instance from the drop down list and then click **Select**.



7. Click on the Project **Overview** tab.



Step 2: Adding a Data Asset to the DSX Labs project

1. Download the Bike Share data file from <https://github.com/bleonardb3/Proof-of-Technology/blob/master/DSX/Lab-1/data/BikeShare.csv>

The data in this file has already been prepared and it ready to be input into the Modeling step.

2. Right click on Raw, and click on Save link as

bleonard3 / Proof-of-Technology

Proof-of-Technology / DSX / Lab-1 / data / BikeShare.csv

729 lines (728 sloc) | 37 KB

Raw Blame History

	Month	Ride_Count	temp avg	dew point avg	humidity % avg	visibility (mi) avg	wind (mph) avg	precip. (in)	Season	Day_of_Week	Fog	Rain	Snow	Thunde
1	1	2611	37	16	45	10	21	0.000000	winter	5	0	0	0	0
2	1	5155	42	24	48	10	12	0.000000	winter	6	0	0	0	0
3	1	1718	38	35	73	4	9	0.500000	winter	7	1	1	0	0
4	1	3711	55	45	68		26	0.200000	winter	1	0	1	0	0
5	1	5262	41	12	35	10	28	0.000000	winter	2	0	0	0	0
6	1	2086	28	16	55	6	14	0.200000	winter	3	0	0	1	0
7	1	3063	23	5	49	10	30		winter	4	0	0	1	0
8	1	2976	19	-4	36	10	24	0.000000	winter	5	0	0	0	0
9	1	4237	34	7	34	10	28	0.000000	winter	6	0	0	0	0
10	1	2694	25	-1	32	10	20	0.000000	winter	7	0	0	0	0

3. Go back to the DSX-Labs project. Click on **add data assets** or the  icon.

IBM Data Science Experience

My Projects > DSX Labs

Overview Analytics Assets Data Assets Bookmarks Deployments Collaborators Settings

Notebooks view all (2)

NAME	SHARED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
Machine Learning with Spark - Joel			Python 3.5	Bernard Beekman	11 Jul 2017	
Connect and Interact with DashDB			Python 3.5	Bernard Beekman	10 Jul 2017	

Data Assets view all (2)

NAME	TYPE	SERVICE	LAST MODIFIED	ACTIONS
Female-human-trafficking.csv	CSV	Object Storage (Swift API)		
myDashDB	Connection	dashDB	5 Jul 2017	

Bookmarks view all (1)

4. Click on browse and then go to the folder where the BikeShare.csv is stored. Select BikeShare.csv and then click Open.

The screenshot shows the IBM Data Science Experience interface. The top navigation bar includes 'Projects', 'Catalog', 'Tools', 'Data Services', and 'Community'. The main content area is divided into sections: 'Notebooks', 'Data Assets', and 'Bookmarks'. The 'Data Assets' section is currently active, displaying a table of assets. A sidebar on the right shows a file upload area and a list of files.

NAME	TYPE	SERVICE	LAST MODIFIED	ACTIONS
BikeShare.csv	CSV	Object Storage (Swift API)		...
Female-human-trafficking.csv	CSV	Object Storage (Swift API)		...
myDashDB	Connection	dashDB	5 Jul 2017	...

Step 3: Create a Model to predict the Ride Count

1. Click on the Analytic Assets Tab

The screenshot shows the IBM Data Science Experience interface with the 'Analytics Assets' tab selected. The 'Data Assets' section is active, displaying a table of assets. A sidebar on the right shows a file upload area and a list of files.

NAME	TYPE	SERVICE	LAST MODIFIED	ACTIONS
BikeShare.csv	CSV	Object Storage (Swift API)		...
Female-human-trafficking.csv	CSV	Object Storage (Swift API)		...
myDashDB	Connection	dashDB	5 Jul 2017	...

2. Click on the **add models**.

The screenshot shows the IBM Data Science Experience dashboard. The top navigation bar includes 'Projects', 'Catalog', 'Tools', 'Data Services', and 'Community'. The main content area is titled 'Analytics Assets' and contains three sections: 'Notebooks', 'Models', and 'Streaming Pipelines'. The 'Models' section has a table with columns: NAME, STATUS, RUNTIME, LAST MODIFIED, and ACTIONS. There are two rows of models listed. A blue arrow points to the '+ add models' button in the top right corner of the 'Models' section.

NAME	STATUS	RUNTIME	LAST MODIFIED	ACTIONS
Female Human Trafficking- Manual	untrained		21 Aug 2017	...
Female Human Trafficking	untrained		9 Jul 2017	...

3. Enter the Model Name, Description, Select Manual, and click on **Create**.

The screenshot shows the 'Create new model' form. It has fields for 'Name' (containing 'Bike Share') and 'Description' (containing 'Create model on the Bike Share data'). Below these are dropdown menus for 'Machine Learning Service' (set to 'Watson Machine Learning') and 'Spark Service' (set to 'Fisheries'). At the bottom, there are two radio buttons: 'Automatic' and 'Manual'. The 'Manual' option is selected and highlighted with a blue arrow. At the bottom right, there are 'Cancel' and 'Create' buttons. A blue arrow points to the 'Create' button.

Name
Bike Share

Description
Create model on the Bike Share data

Machine Learning Service
Watson Machine Learning

Spark Service
Fisheries

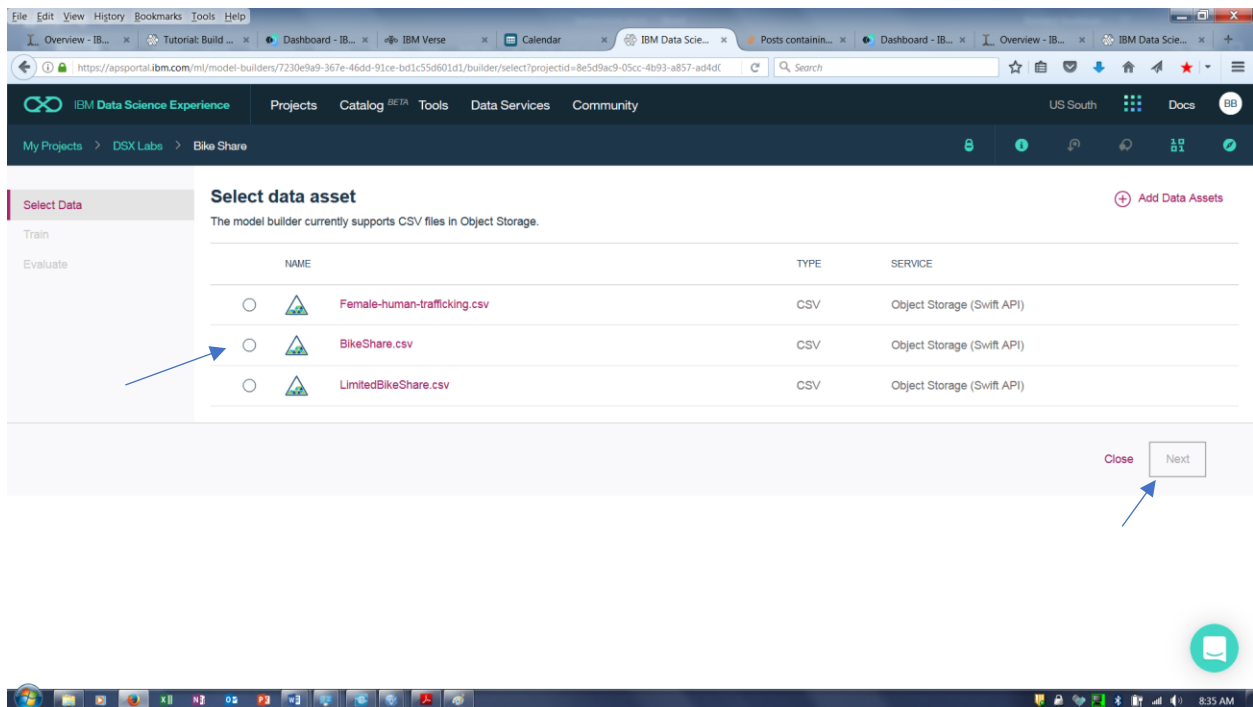
Automatic
Prepare my data and create a model automatically

Manual
Let me prepare my data and select which models to train

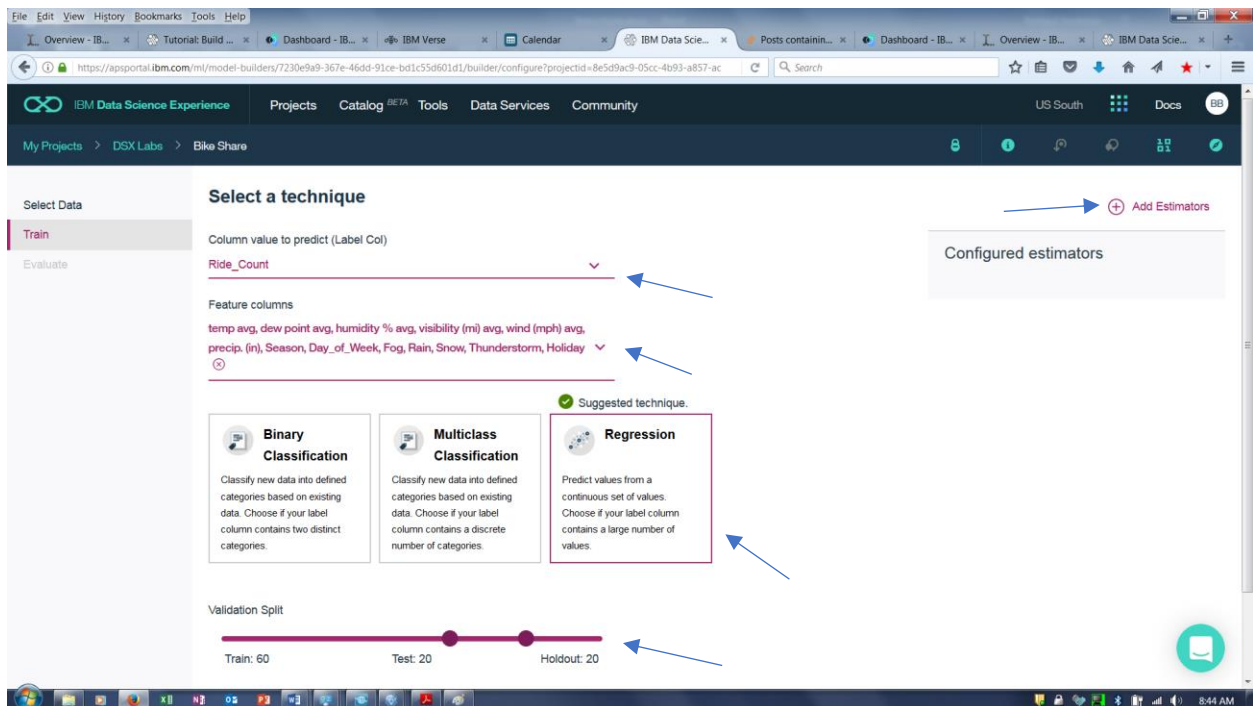
Need something more flexible? Create a [notebook](#) or design a [flow](#).

Cancel Create

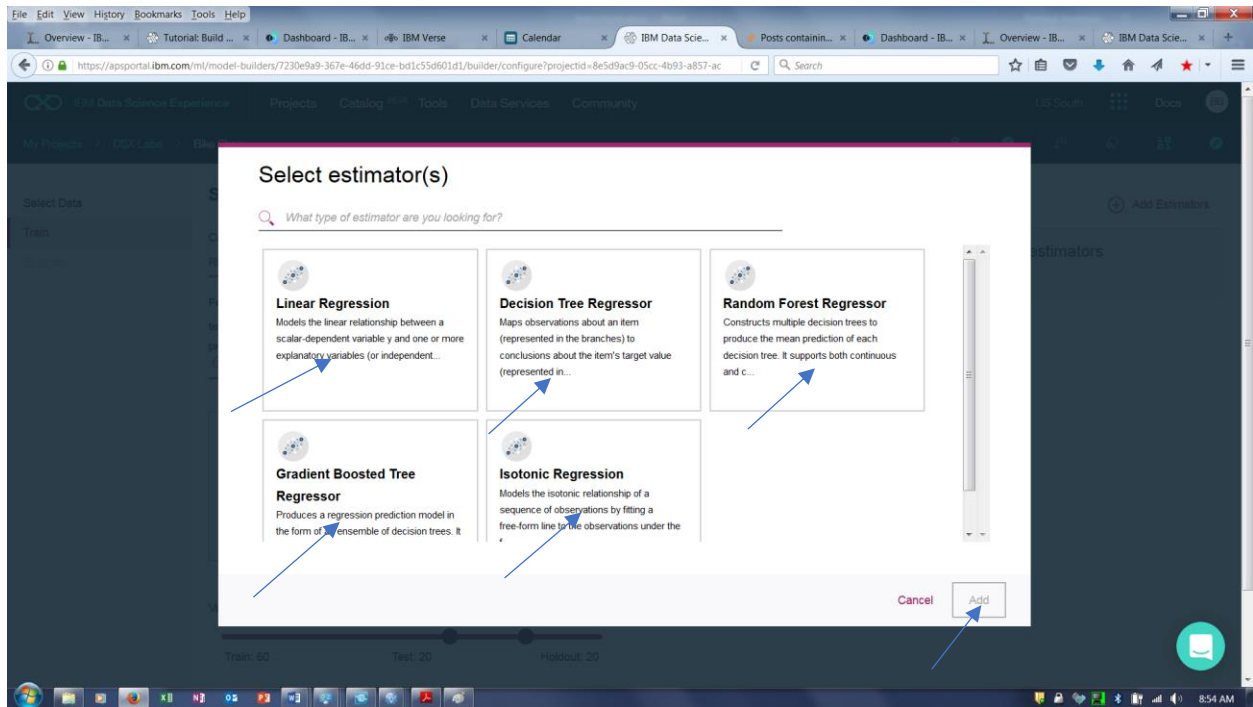
4. Click on the **BikeShare.csv** and click on **Next**



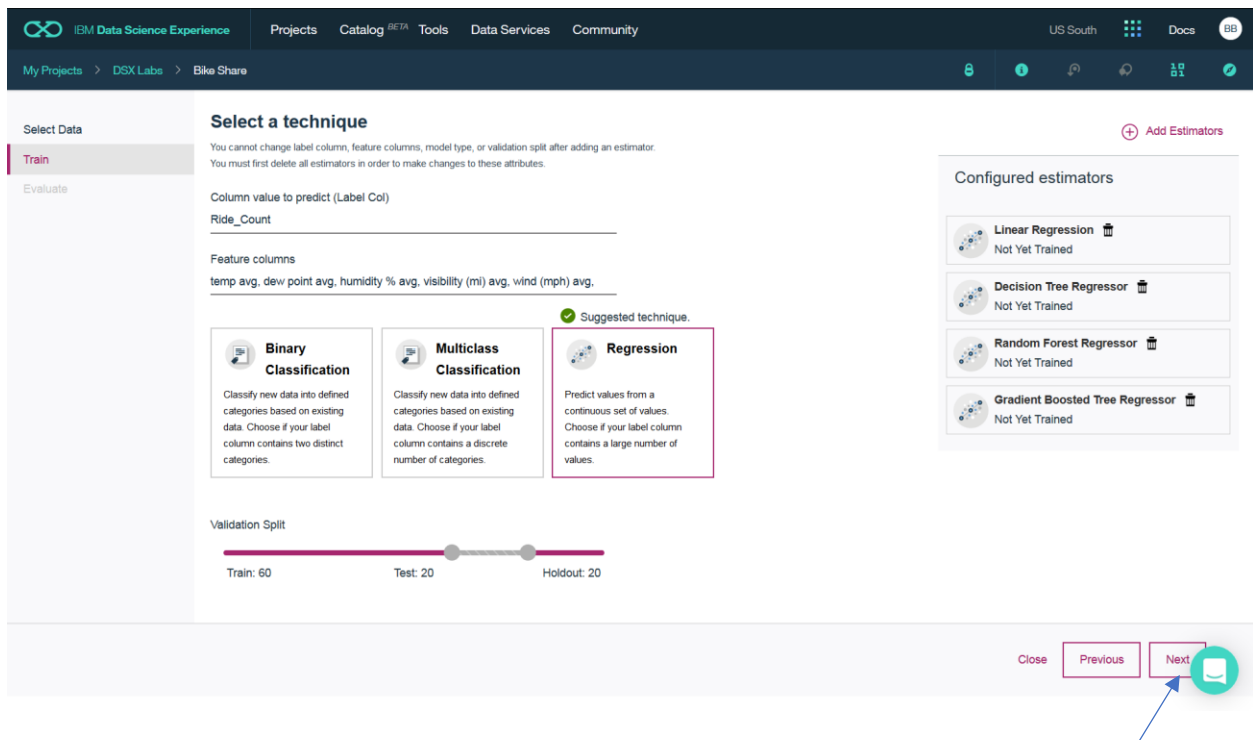
- For **Column value to predict (Label Col)** select Ride_Count. For **Feature columns** select the features that you want to include. Click on the **Regression** Box (which is suggested by the service). Adjust the **Validation Split** as desired. Click on **Add Estimators** to add the specific models to use.



6. Select the specific estimators – you can select any or all of them. Select all estimators. Select **Add**.



7. Select the **Next** button.



8. The system trains and evaluates each model. The models are listed in ascending order of quality with the best result at the bottom. Click on the **RandomForestRegression** and then click **Save**.

The screenshot shows the IBM Data Science Experience interface. On the left, the 'Evaluate' tab is selected. The main area displays a table titled 'Select model' with the following data:

	ESTIMATOR TYPE	STATUS	PERFORMANCE	ROOT MEAN SQUARED ERROR	MEAN SQUARED ERROR	R2	EXPLAINED VARIANCE	MEAN ABSOLUTE ERROR	LAST EVALUATION	ACTIONS
<input type="radio"/>	IsotonicRegression	Trained & Evaluated	Excellent	2330.65048	5431931.65294	0.56448	7612583.07126	1719.59569	21 Aug 2017, 9:09 AM	...
<input type="radio"/>	DecisionTreeRegressor	Trained & Evaluated	Excellent	2133.16918	4550410.76059	0.63516	11536943.50087	1516.43649	21 Aug 2017, 9:07 AM	...
<input type="radio"/>	GBTRegressor	Trained & Evaluated	Excellent	2076.33825	4311180.514	0.65434	11644366.54494	1434.48909	21 Aug 2017, 9:08 AM	...
<input type="radio"/>	LinearRegression	Trained & Evaluated	Excellent	1925.59957	3707933.69559	0.70271	9666427.995	1448.8493	21 Aug 2017, 9:07 AM	...
<input checked="" type="radio"/>	RandomForestRegressor	Trained & Evaluated	Excellent	1796.78288	3228428.71119	0.74115	8061082.34624	1272.43294	21 Aug 2017, 9:07 AM	...

At the bottom right, there are buttons for 'Close', 'Previous', and 'Save'. The 'Save' button is highlighted with a blue arrow.

9. The system displays the model training summary. To run a sample prediction, select the **Predictions** tab

The screenshot shows the IBM Data Science Experience interface. The 'Predictions' tab is selected. The main area displays the 'Bike Share' project details. The 'Predictions' tab is active, showing the following information:

Machine learning service	Watson Machine Learning
Label column	Ride_Count
Algorithm	org.apache.spark.ml.regression.RandomForestRegressionModel
Model builder details	View
Training data schema	View
Input data schema	View
Runtime environment	spark-2.0
Training date	21 Aug 2017, 9:23 AM

Below the table, there is a section for 'Deployments'. It shows a table with columns 'NAME', 'DEPLOYMENT TYPE', and 'ACTIONS'. The message 'Your model is not deployed.' is displayed.

10. Enter values for the input features and then click on **Predict**.

The screenshot shows the IBM Data Science Experience web interface. The top navigation bar includes 'Projects', 'Catalog', 'Tools', 'Data Services', and 'Community'. The left sidebar shows 'My Projects' with 'DSX Labs' and 'Bike Share'. The main content area is titled 'Spark Service' and 'Fisheries'. Under 'Prediction input data', there are input fields for 'Month' (1), 'temp avg' (37), 'dew point avg' (16), 'humidity % avg' (45), and 'visibility (mi) avg'. A 'Predict' button is located at the bottom left of the input section. The right side of the interface is a large, empty light blue area.

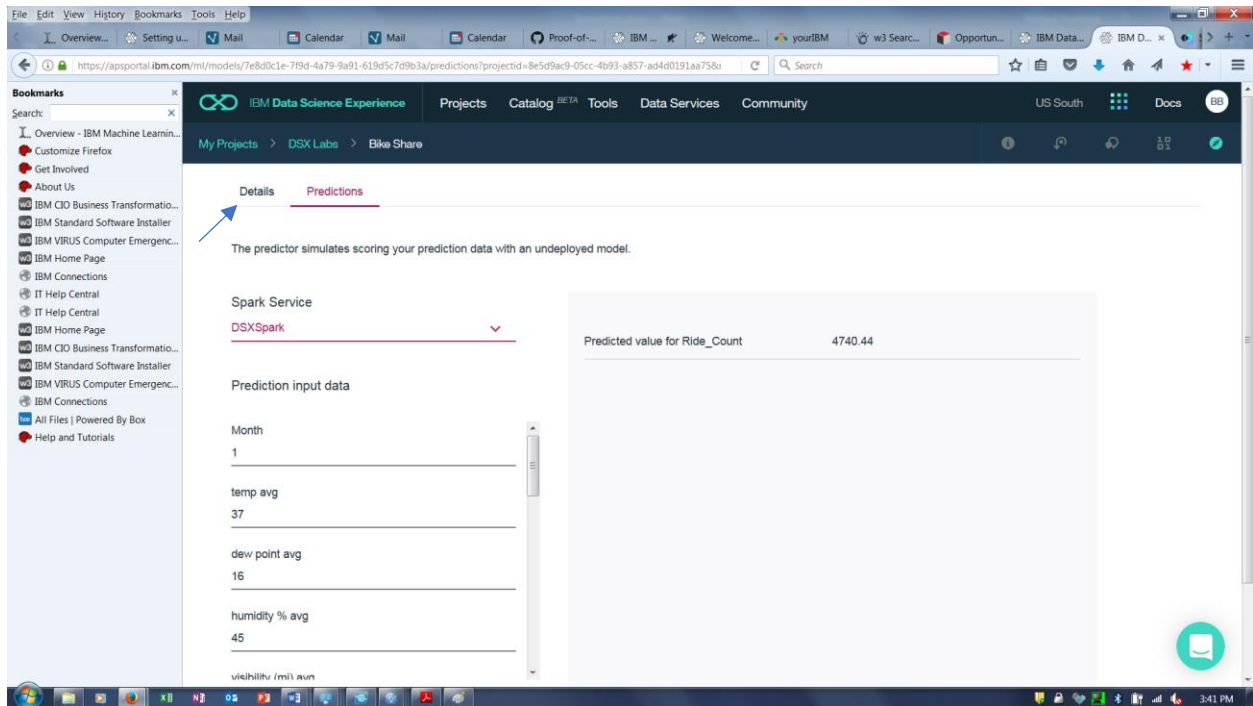
11. The estimated Ride_Count is then displayed

The screenshot shows the same IBM Data Science Experience interface as before, but now the 'Predicted value for Ride_Count' is displayed as 4352.03. A blue arrow points to this value. The input fields and 'Predict' button remain the same on the left side.

Step 4: Deploying a Model

We can deploy the model to enable applications to invoke it via an API call.

1. Select the **Details** Tab



2. Scroll down to the **Add Deployments** option. Click on **Add Deployments**

The screenshot shows the IBM Data Science Experience interface. On the left is a sidebar with a 'Bookmarks' section and a search bar. The main area displays the 'Bike Share' project details under the 'Predictions' tab. A table lists project metadata:

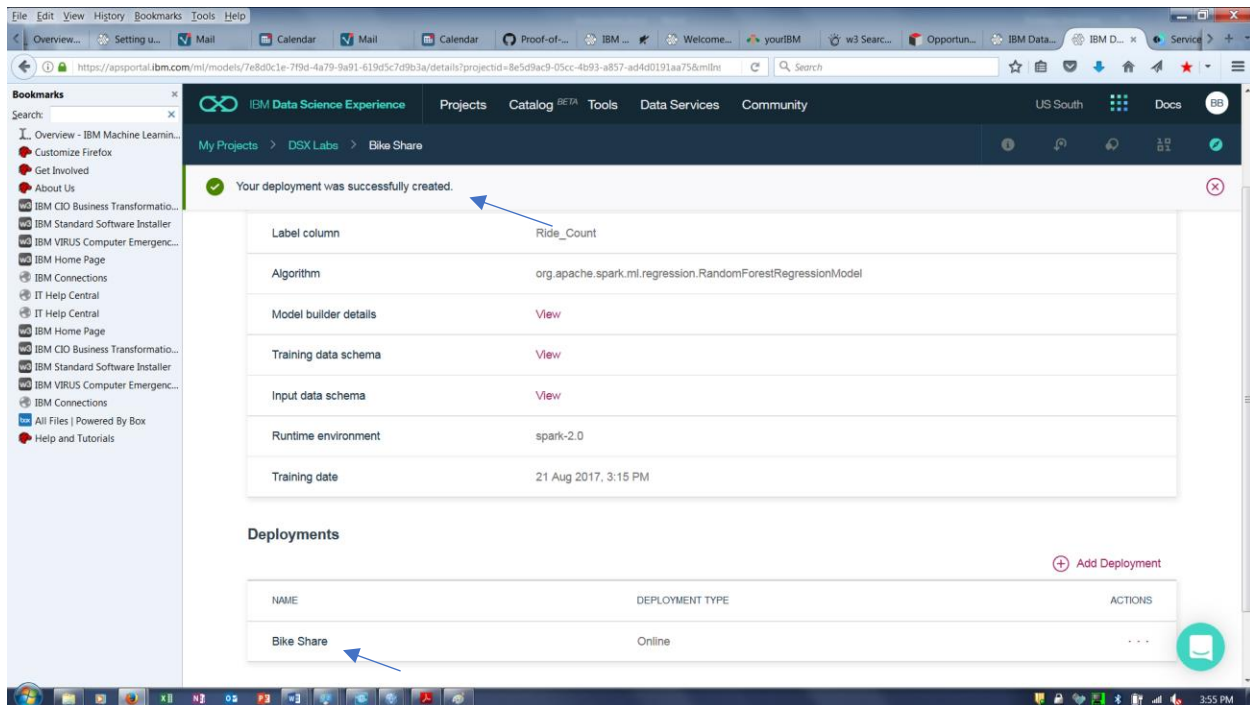
Property	Value
Machine learning service	Watson Machine Learning
Label column	Ride_Count
Algorithm	org.apache.spark.ml.regression.RandomForestRegressionModel
Model builder details	View
Training data schema	View
Input data schema	View
Runtime environment	spark-2.0
Training date	21 Aug 2017, 3:15 PM

Below the table is the 'Deployments' section, which is currently empty. A message states 'Your model is not deployed.' An arrow points to the 'Add Deployment' button in the top right corner of the deployments table.

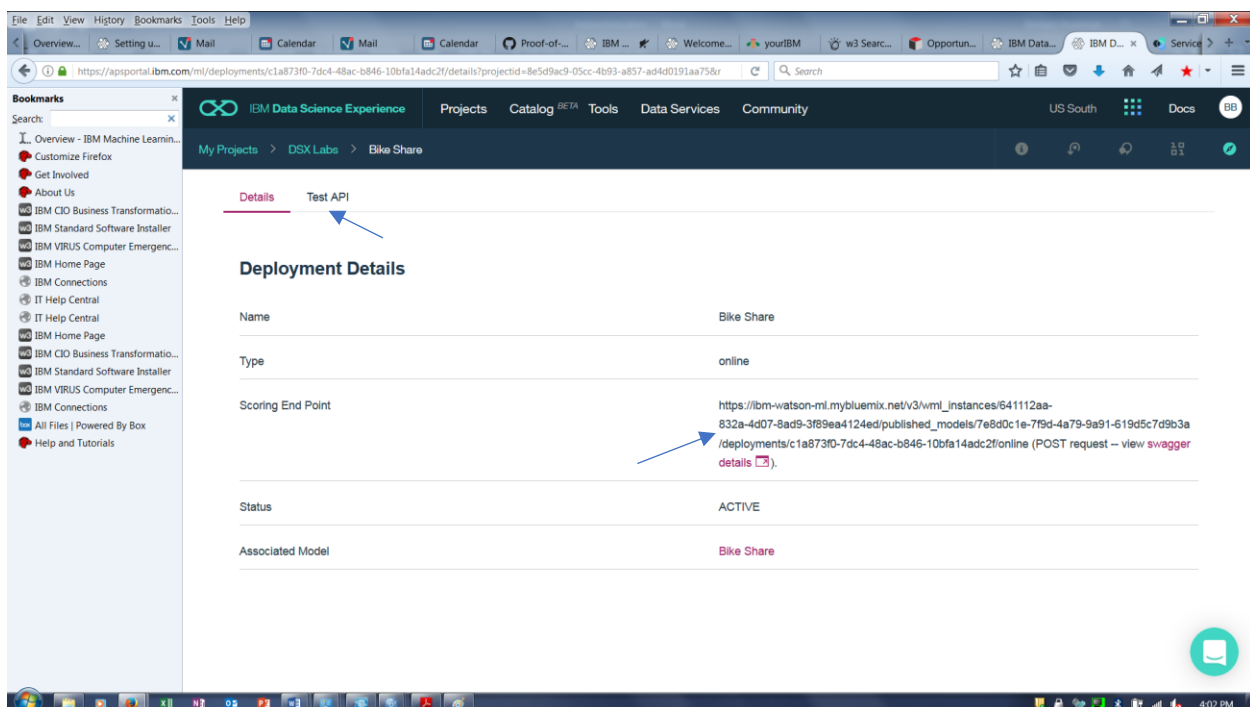
3. Select Online for **Deployment Type**, enter Bike Share for **Name**, and click on **Deploy**.

This screenshot shows the 'Deploy model' dialog box overlaid on the project details page. The dialog has two input fields: 'Deployment Type' and 'Name'. The 'Deployment Type' dropdown is set to 'Online', and the 'Name' field contains 'Bike Share'. At the bottom right of the dialog, there are 'Close' and 'Deploy' buttons. An arrow points to the 'Deploy' button. The background shows the same 'Bike Share' project details as the previous screenshot.

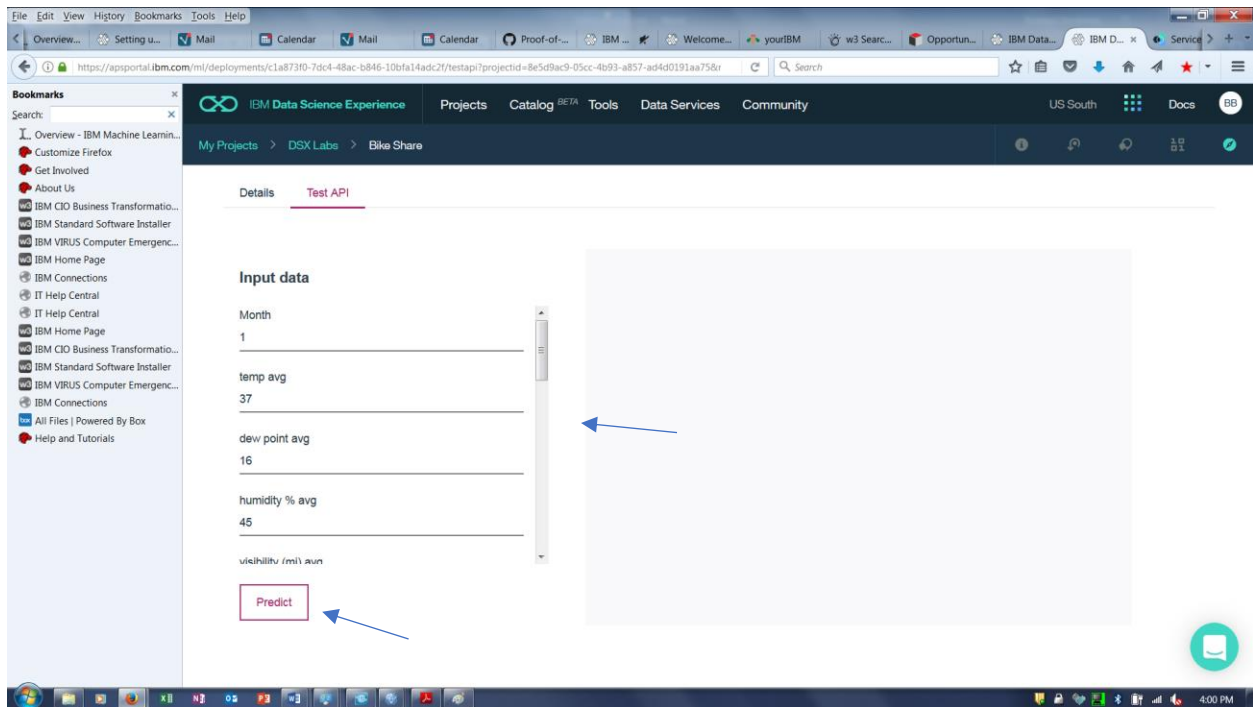
4. The system responds with an acknowledgement that the model was successfully deployed. Click on **Bike Share** to test the deployed API.



- The system displays information about the deployed service include the endpoint to invoke. Click on **Test API** to test out the API.



- Enter values for the input parameters and then click on **Predict**.



7. The predicted result is returned.

