

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
5. Creating a web app to invoke the Model.

Setting up your environment

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

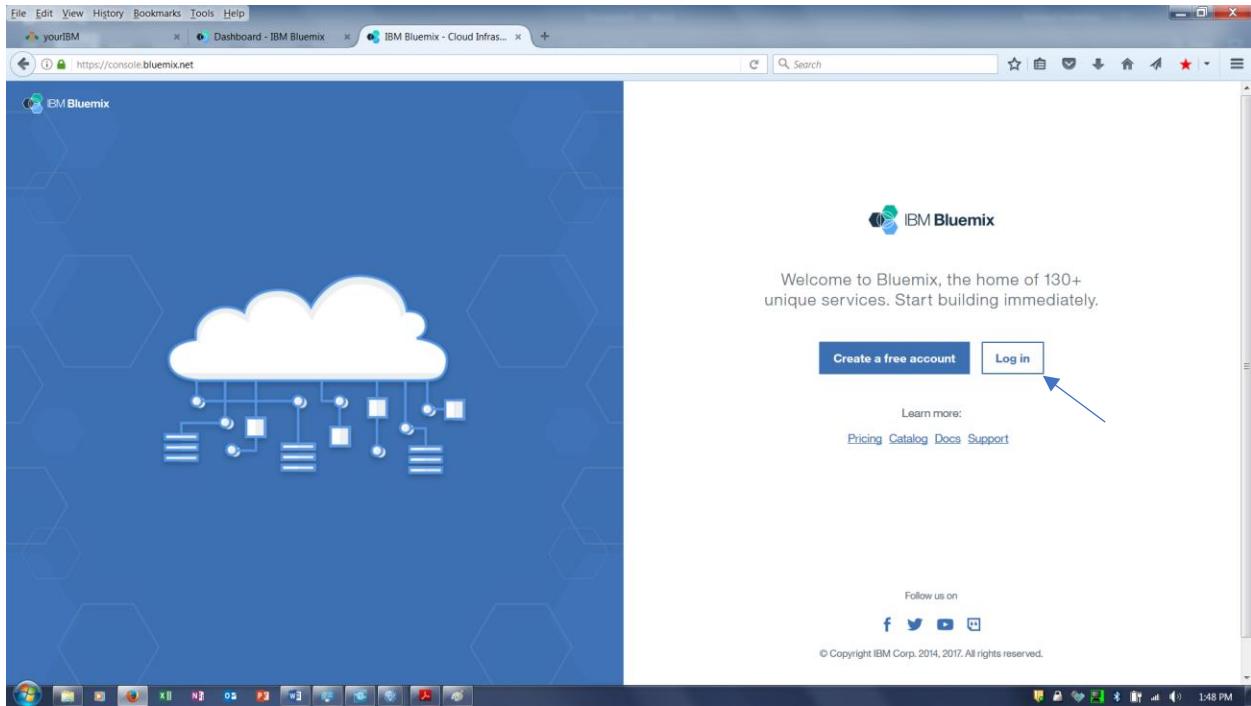
- Watson Machine Learning
- Object Storage
- Apache Spark

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.

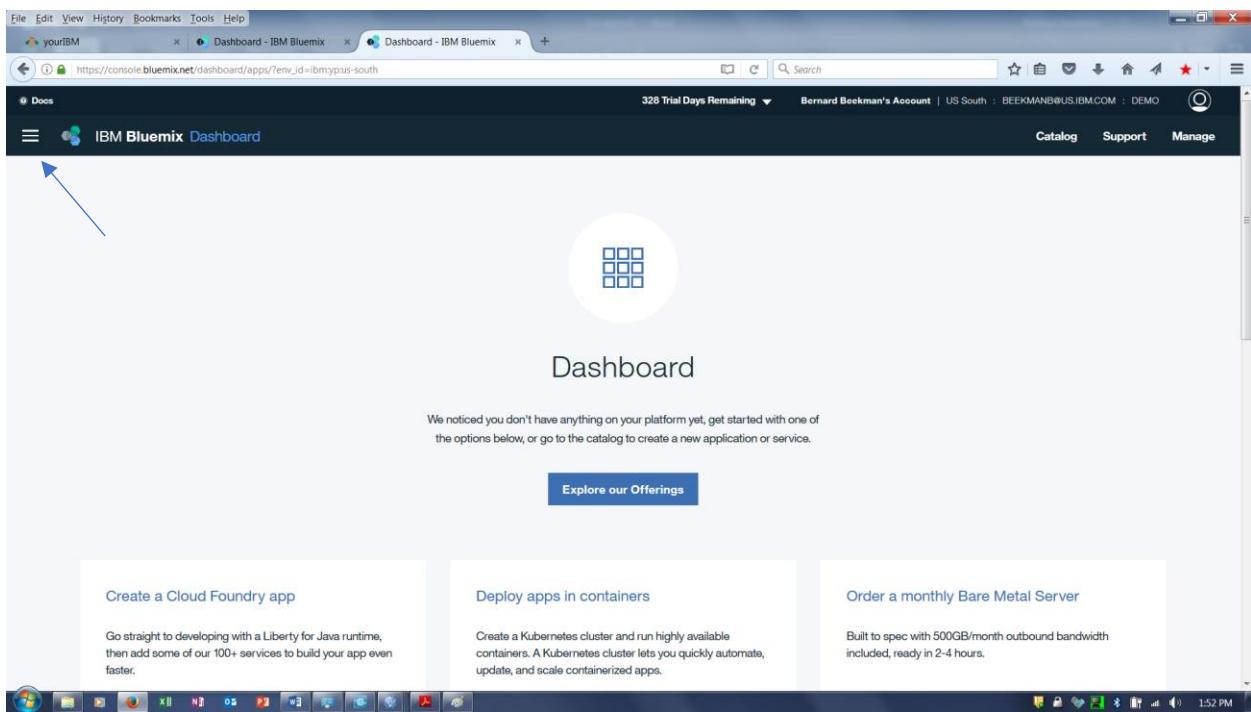
Creating a Machine Learning Instance

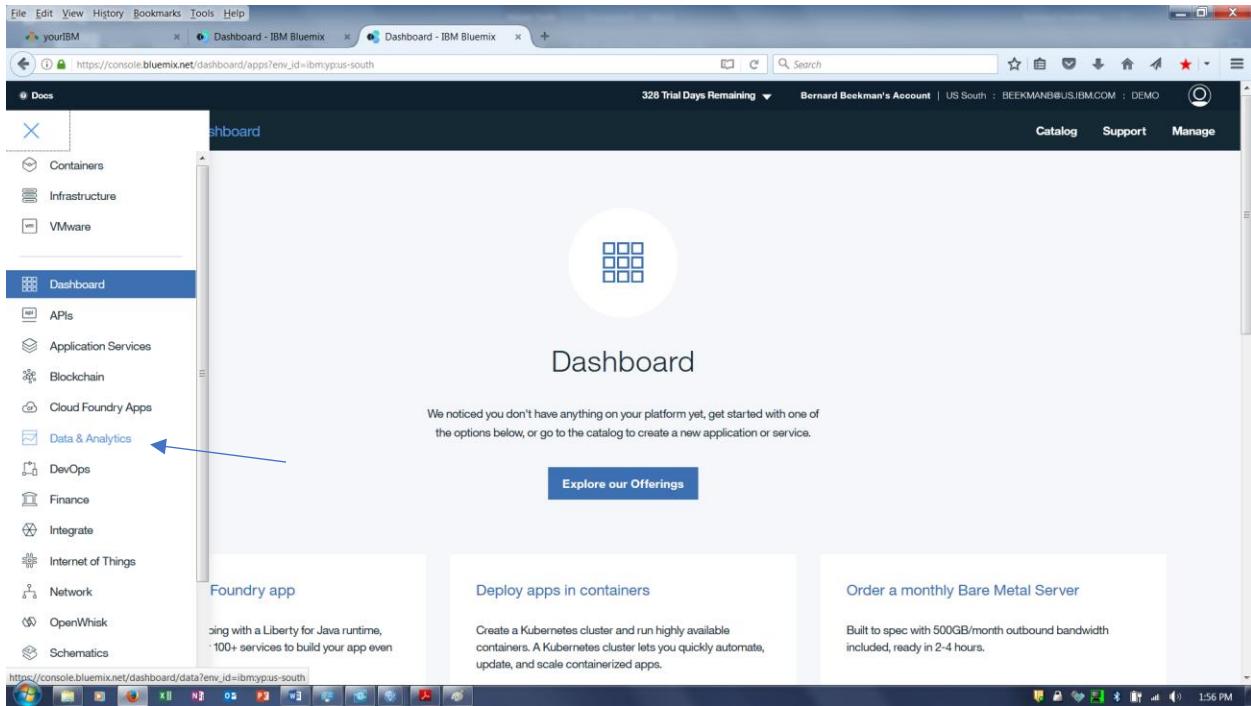
To create a Machine Learning Instance, you must perform the following steps:

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



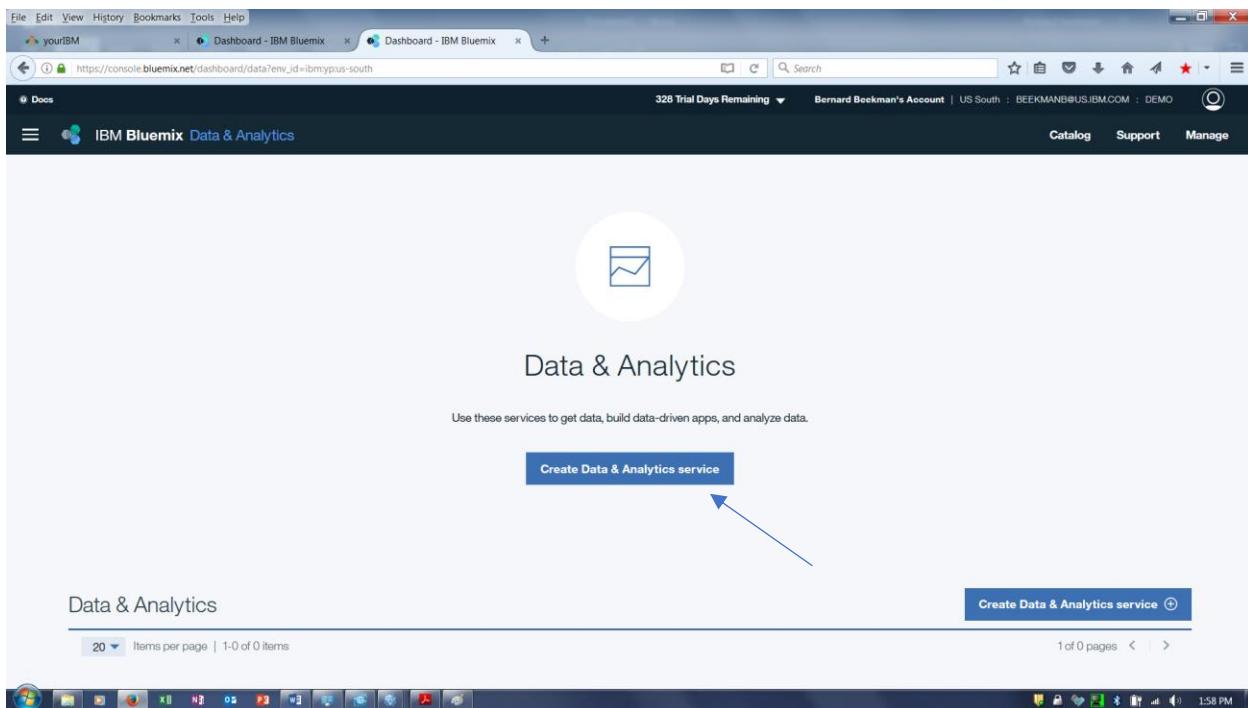
- Once logged in, click on the hamburger icon, and from the navigation panel, click **Data & Analytics**.



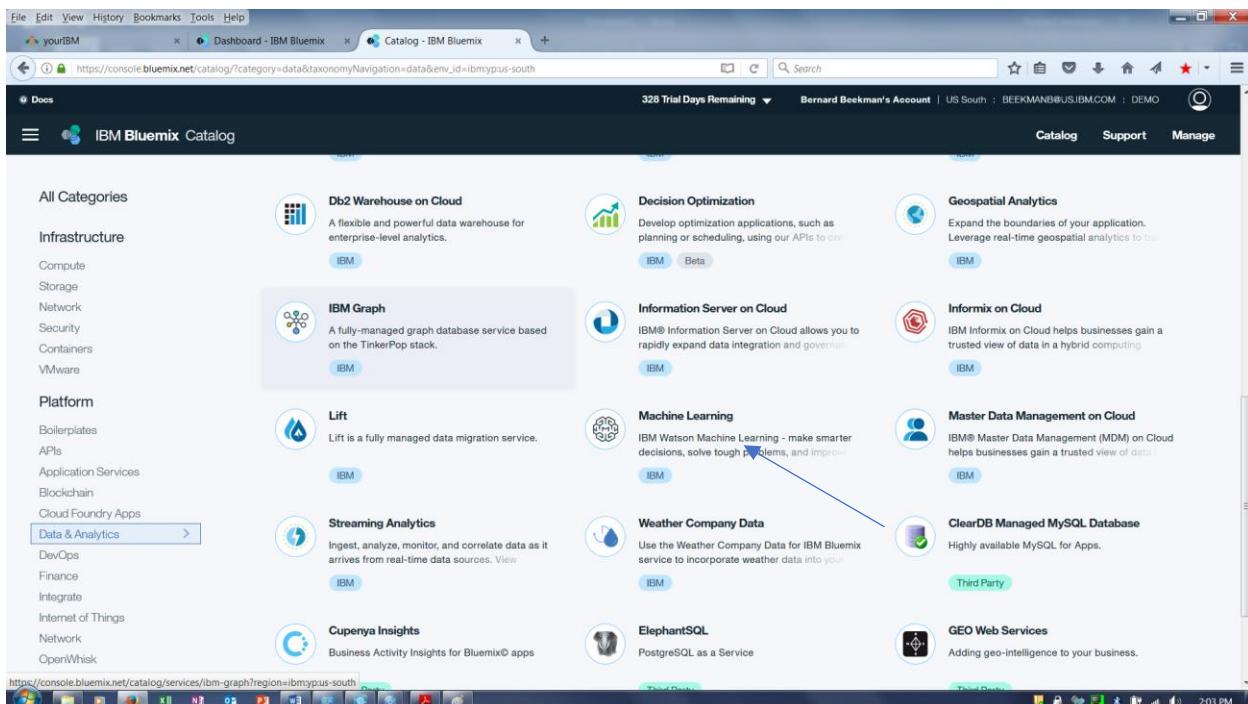


You see a screen centered on data services. You can return here regularly to work with your data and analytics services from one easy-to-use page. Check to see if a Machine Learning service already exists. If not, continue, otherwise go to **Adding existing Bluemix instances to a project in Data Science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

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



4. Scroll down to Machine Learning and click.



5. Configure service.

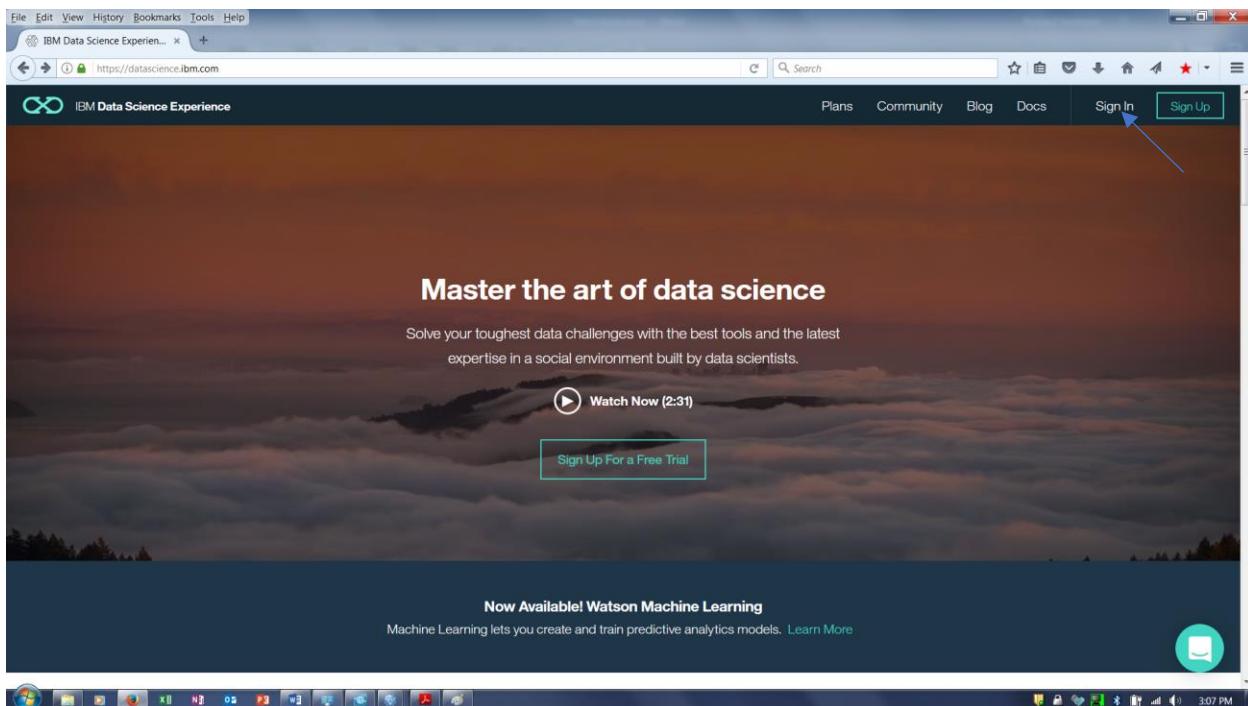
Enter a descriptive name for your service, choose a space, and select your data plan (find plan comparison and pricing details on this page). Click on **Create**.

The screenshot shows the IBM Bluemix Catalog interface. A blue arrow points from the 'Service name' field ('Machine Learning') to the 'Create' button at the bottom right. Another blue arrow points from the 'Select region to deploy in' dropdown ('US South') to the same 'Create' button. A third blue arrow points from the 'Choose a space:' dropdown ('DEMO') to the 'Create' button. A fourth blue arrow points from the 'Connect to:' dropdown ('Leave unbound') to the 'Create' button. The 'Service name:' field contains 'Machine Learning'. The 'Credential name:' field contains 'Credentials-1'. The 'Select region to deploy in:' dropdown is set to 'US South'. The 'Choose an organization:' dropdown contains 'BEEKMANB@US.IBM.COM'. The 'Choose a space:' dropdown is set to 'DEMO'. The 'Connect to:' dropdown is set to 'Leave unbound'. At the bottom right is a large blue 'Create' button.

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.

Name	Role	Collaborators	Date Created	Last Updated
GBS Bootcamp	Admin	BB	Aug 19, 2017	Aug 20, 2017
DSX Lab	Admin	BB	Mar 28, 2017	Aug 19, 2017
Machine Learning	Admin	BB, SP, JP, +3	Aug 08, 2017	Aug 18, 2017

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

The screenshot shows the IBM Data Catalog interface. At the top, there are several tabs: 'IBM Data Catalog' (selected), 'Overview - IBM Machine L...', 'Tutorial: Build a logistic re...', and 'Mail'. Below the tabs, the main navigation bar includes 'Projects', 'Catalog BETA', 'Governance BETA', and 'Data Services'. The 'Projects' tab is selected, showing the 'My Projects' section. A search bar at the top says 'Find in my projects' and a dropdown menu says 'All Projects'. A button to 'create project' is also present. The main area displays a table of projects:

NAME	ROLE	COLLABORATORS	CREATOR	LAST MODIFIED	ACTIONS
GBS Bootcamp	Admin		Bernard Beekman	20 Aug 2017	...
DSX Lab	Admin		Bernard Beekman	19 Aug 2017	...
Machine Learning	Admin		Joel Patterson	18 Aug 2017	...
DSX Labs	Admin		Bernard Beekman	9 Jul 2017	...
Education Analysis	Admin		Bernard Beekman	6 Jun 2017	...
Ed Analytics Framework	Admin		Erica Layne Morrison	25 Apr 2017	...
Hackathon	Admin		Bernard Beekman	7 Apr 2017	...
GWU-Test	Admin		Bernard Beekman	7 Apr 2017	...
GWU	Admin		Bernard Beekman	7 Apr 2017	...
Default project	Admin		Bernard Beekman	7 Apr 2017	...

4. Select the Settings Tab.

The screenshot shows the 'DSX Labs' project details. The navigation bar at the top includes 'Overview', 'Analytics Assets', 'Data Assets', 'Bookmarks', 'Deployments', 'Collaborators', and the 'Settings' tab, which is highlighted. Below the navigation bar, there are three sections: 'Notebooks', 'Data Assets', and 'Bookmarks'.

- Notebooks:** Shows two entries: 'Machine Learning with Spark - Joel' and 'Connect and Interact with DashDB'. An 'add notebooks' button is available.
- Data Assets:** Shows two entries: 'Female-human-trafficking.csv' (CSV, Object Storage (Swift API)) and 'myDashDB' (Connection, dashDB). An 'add data assets' button is available.
- Bookmarks:** Shows one entry: 'A comparison of logistic regression a...'. An 'explore community' button is available.

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

The screenshot shows the IBM Data Catalog interface. At the top, there are several tabs: 'IBM Data Catalog', 'Overview - IBM Machine L...', 'Tutorial: Build a logistic re...', and 'Mail'. Below the tabs, the main navigation bar includes 'Projects', 'Catalog BETA', 'Governance BETA', and 'Data Services'. The 'My Projects' section shows 'DSX Labs'. On the right, there are icons for 'US South', 'Docs', and 'BB'. The main content area has a heading 'Storage' with a sub-section for 'Object Storage (Swift API)'. It shows '338.03 KB Used' and '0.0% of 5GB Used'. Below this is a section titled 'Associated Services' with a table:

NAME	SERVICE TYPE	PLAN
Fisheries	Spark	Personal

A dropdown menu is open next to the 'Machine Learning' service instance, listing options like 'Amazon EMR Spark', 'IBM Analytics Engine', 'Machine Learning', and 'Spark'. A blue arrow points from the text 'Select the Machine Learning service instance from the drop down list and then click Select.' to this dropdown menu.

Associated Services

NAME	SERVICE TYPE	PLAN
Fisheries	Spark	Personal

Access Tokens

NAME	ROLE	CREATED	LAST USED	ACTIONS
you currently have no access tokens				

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

The screenshot shows the IBM Data Science Experience interface. At the top, there are several tabs: 'Overview - IBM Machine L...', 'Tutorial: Build a logistic re...', 'Dashboard - IBM Bluemix', and 'IBM Data Science Experience...'. Below the tabs, the main navigation bar includes 'Projects', 'Catalog BETA', 'Tools', 'Data Services', and 'Community'. The 'My Projects' section shows 'DSX Labs'. On the right, there are icons for 'US South', 'Docs', and 'BB'. The main content area has a heading 'Add Machine Learning Service' with a sub-section for 'Existing Service Instance'. It shows 'Watson Machine Learning' in the list. There are 'Select' and 'Cancel' buttons at the bottom of the dialog.

Add Machine Learning Service

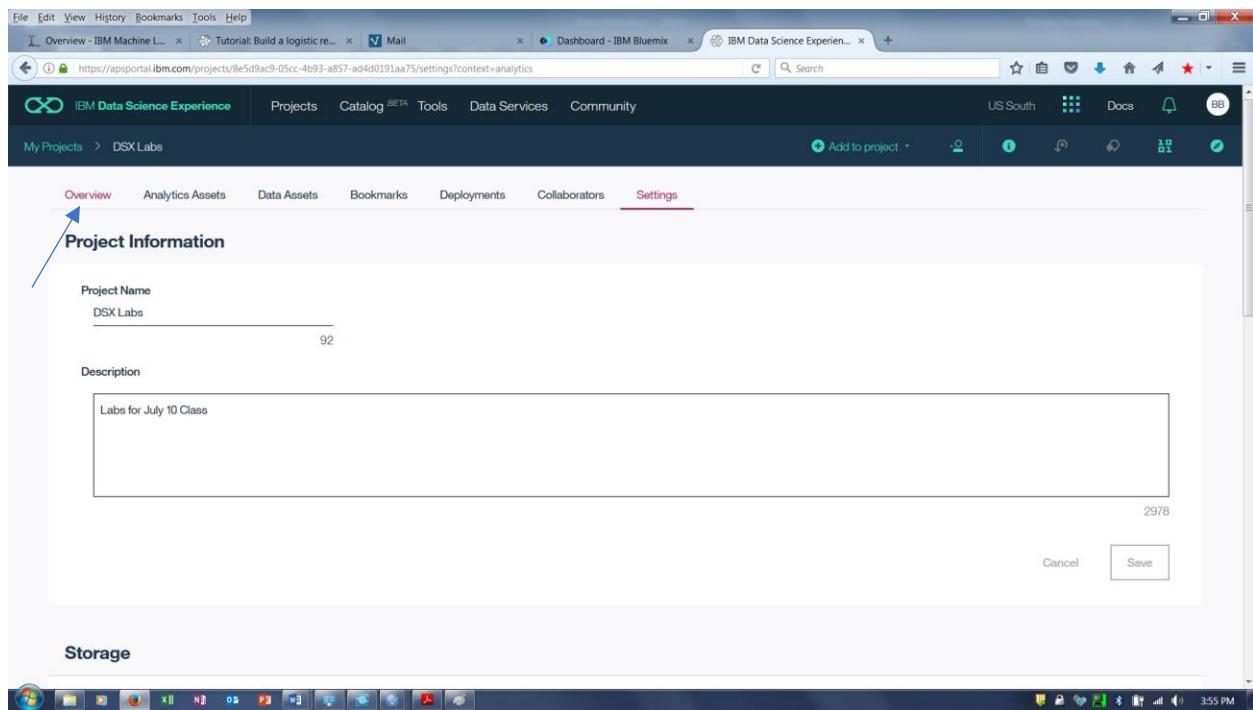
Existing New

Existing Service Instance

Watson Machine Learning

Select Cancel

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



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

BikeShare.csv														
	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	Thunder
1	2	2611	37	16	45	10	21	0.000000	winter	5	0	0	0	0
3	1	5155	42	24	48	10	12	0.000000	winter	6	0	0	0	0
4	1	1718	38	35	73	4	9	0.500000	winter	7	1	1	0	0
5	1	3711	55	45	68		26	0.200000	winter	1	0	1	0	0
6	1	5262	41	12	35	10	28	0.000000	winter	2	0	0	0	0
7	1	2086	28	16	55	6	14	0.200000	winter	3	0	0	1	0
8	1	3063	23	5	49	10	30		winter	4	0	0	1	0
9	1	2976	19	-4	36	10	24	0.000000	winter	5	0	0	0	0
10	1	4237	34	7	34	10	28	0.000000	winter	6	0	0	0	0
11	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.

The screenshot shows the DSX interface with the following sections:

- Overview**: Shows two notebooks: "Machine Learning with Spark - Joel" and "Connect and Interact with DashDB".
- Analytics Assets**: Shows two notebooks: "Machine Learning with Spark - Joel" and "Connect and Interact with DashDB".
- Data Assets**: Shows a single data asset named "Female-human-trafficking.csv" (Type: CSV, Service: Object Storage (Swift API)).
- Bookmarks**: Shows one bookmarked article titled "A comparison of logistic regression a...".

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 main navigation bar includes 'File', 'Edit', 'View', 'History', 'Bookmarks', 'Tools', and 'Help'. Below the bar, there are several tabs: 'Overview - IBM Machi...', 'Tutorial: Build a logist...', 'Mail', 'Dashboard - IBM Blue...', 'IBM Data Science Exp...', 'Copy file from github ...', 'Proof-of-Technology/...', and 'Proof-of-Technology/...'. The current view is under 'My Projects > DSX Labs'. The main content area has tabs for 'Overview', 'Analytics Assets', 'Data Assets', 'Bookmarks', 'Deployments', 'Collaborators', and 'Settings'. The 'Analytics Assets' tab is selected, showing two sections: 'Notebooks' (Machine Learning with Spark - Joel, Connect and Interact with DashDB) and 'Data Assets' (BikeShare.csv, Female-human-trafficking.csv, myDashDB). A sidebar on the right contains 'Files' and 'Connections' sections, with a 'Drop file here or browse' area and a 'Find in Storage' search bar. Two CSV files are listed: 'Female-human-trafficking.csv' and 'BikeShare.csv'. A blue arrow points from the text 'Select the Analytics Assets tab' to the 'Analytics Assets' tab in the interface.

Create a Model to predict the Ride Count

1. Click on the Analytic Assets Tab

The screenshot shows the same IBM Data Science Experience interface as the previous one, but with a blue arrow pointing specifically to the 'Analytics Assets' tab in the top navigation bar. The rest of the interface remains identical, showing the 'DSX Labs' project overview with the 'Analytics Assets' tab selected.

2. Click on the add models.

The screenshot shows the IBM Data Science Experience web interface. The top navigation bar includes File, Edit, View, History, Bookmarks, Tools, and Help. Below the bar, there are several tabs: Overview, Tutorial Build..., Dashboard - IB..., IBM Verse, Calendar, IBM Data Scie..., Posts containin..., Dashboard - IB..., Overview - IB..., and IBM Data Scie... . The main content area is titled 'My Projects > DSX Labs'. It features three sections: 'Notebooks', 'Models', and 'Streaming Pipelines'. The 'Notebooks' section lists two entries: 'Machine Learning with Spark - Joel' and 'Connect and Interact with DashDB'. The 'Models' section lists two entries: 'Female Human Trafficking- Manual' and 'Female Human Trafficking'. The 'Streaming Pipelines' section has a 'add streaming pipelines' button. At the bottom right of the main content area, there is a 'add models' button with a plus sign. A blue arrow points from the 'add models' button in the 'Models' section to the 'add models' button in the top right corner.

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

The screenshot shows the 'Create new model' dialog box. The title bar says 'Create new model BETA'. The form fields include 'Name' (set to 'Bike Share') and 'Description' (set to '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'). There are two options for creating a model: 'Automatic' (which says 'Prepare my data and create a model automatically') and 'Manual' (which says 'Let me prepare my data and select which models to train'). The 'Manual' option is highlighted with a pink box and a blue arrow pointing to it. At the bottom of the dialog, there is a note: 'Need something more flexible? Create a notebook or design a flow.' On the right side, there are 'Cancel' and 'Create' buttons, with the 'Create' button also highlighted with a pink box and a blue arrow pointing to it.

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

The screenshot shows the 'Select data asset' step in the IBM Data Science Experience interface. The left sidebar has 'Train' selected under 'Select Data'. The main area lists three CSV files:

NAME	TYPE	SERVICE
Female-human-trafficking.csv	CSV	Object Storage (Swift API)
BikeShare.csv	CSV	Object Storage (Swift API)
LimitedBikeShare.csv	CSV	Object Storage (Swift API)

A blue arrow points to the 'BikeShare.csv' row. In the bottom right corner, there are 'Close' and 'Next' buttons, with a blue arrow pointing to the 'Next' button.

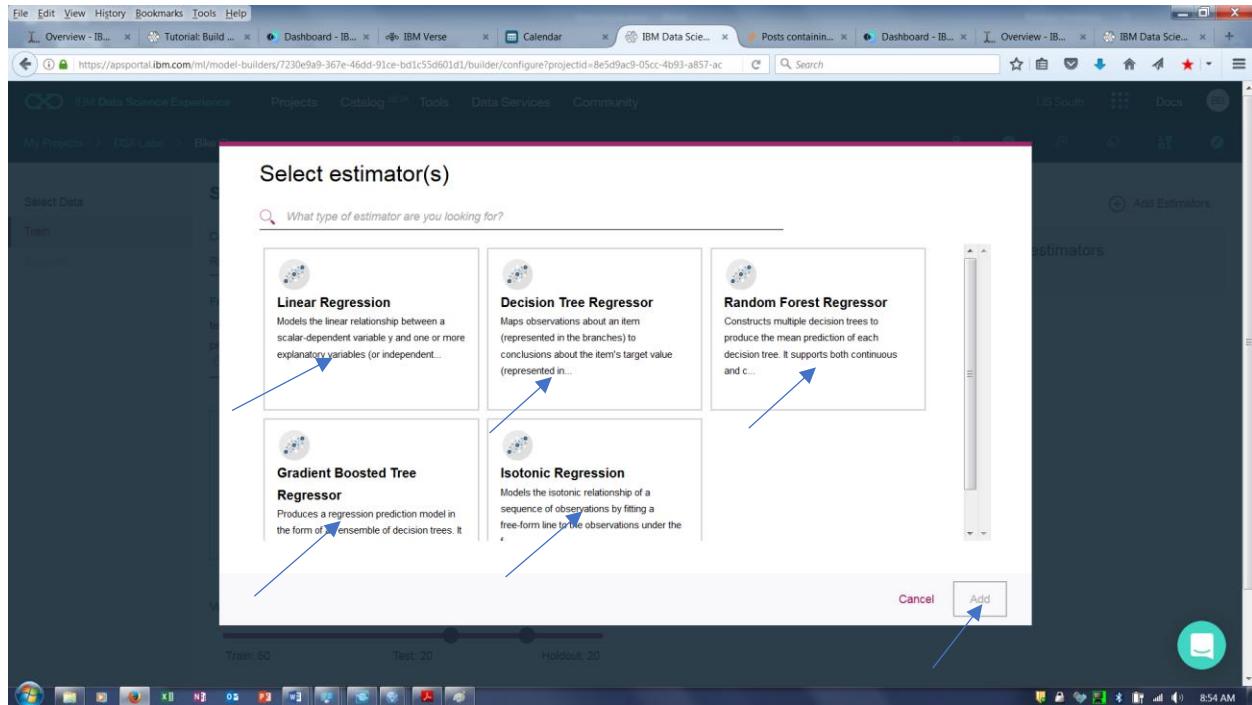
- For **Column value to predict (Label Col)** select Ride_Count. For **Feature columns** select the features that you want to include. Click on the **Regression Box** (which is suggested by the service). Adjust the **Validation Split** as desired. Click on **Add Estimators** to add the specific models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

The screenshot shows the 'Select a technique' step in the IBM Data Science Experience interface. The left sidebar has 'Train' selected under 'Select Data'. The main area shows the configuration for the 'Bike Share' project:

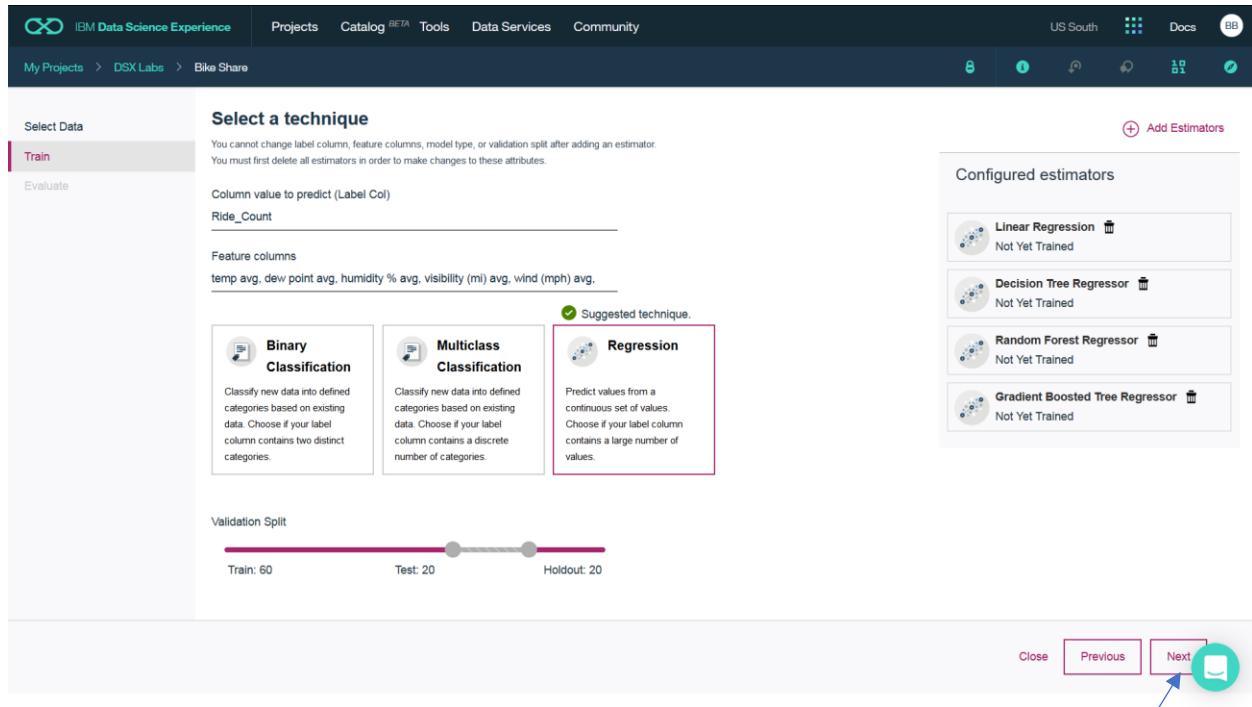
- Column value to predict (Label Col):** Ride_Count
- Feature columns:** temp avg, dew point avg, humidity % avg, visibility (mi) avg, wind (mph) avg, precip. (in), Season, Day_of_Week, Fog, Rain, Snow, Thunderstorm, Holiday
- Suggested technique:** Regression
- Validation Split:** Train: 60, Test: 20, Holdout: 20

Arrows point from the text descriptions to their corresponding UI elements: 'Ride_Count' to the label column dropdown, 'Feature columns' to the dropdown menu, 'Regression' to the highlighted 'Suggested technique' box, and 'Validation Split' to the slider.

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 'Select model' page in the IBM Data Science Experience. On the left, there's a sidebar with 'Evaluate' selected. The main area displays a table of models:

	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	9666427.996	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	...

In the bottom right corner of the table, there are buttons for 'Close', 'Previous', 'Save', and a message icon.

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

The screenshot shows the 'Predictions' tab in the IBM Data Science Experience. At the top, there are tabs for 'Details' and 'Predictions', with 'Details' currently selected. Below the tabs, there's a section for 'Bike Share' with a trash icon. The main area displays a table of model details:

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 this is a section for 'Deployments' with a table:

Deployments		
+ Add Deployment		
NAME	DEPLOYMENT TYPE	ACTIONS
Your model is not deployed.		

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

The screenshot shows the IBM Data Science Experience interface. The top navigation bar includes 'Projects', 'Catalog BETA', 'Tools', 'Data Services', and 'Community'. A 'US South' dropdown and a 'Docs' button are also present. The main content area is titled 'Spark Service' under 'Fisheries'. On the left, there's a 'Prediction input data' section with fields for 'Month' (1), 'temp avg' (37), 'dew point avg' (16), 'humidity % avg' (45), and 'visibility (mi) avg'. Below these fields is a 'Predict' button. To the right of the input fields is a large, empty white space.

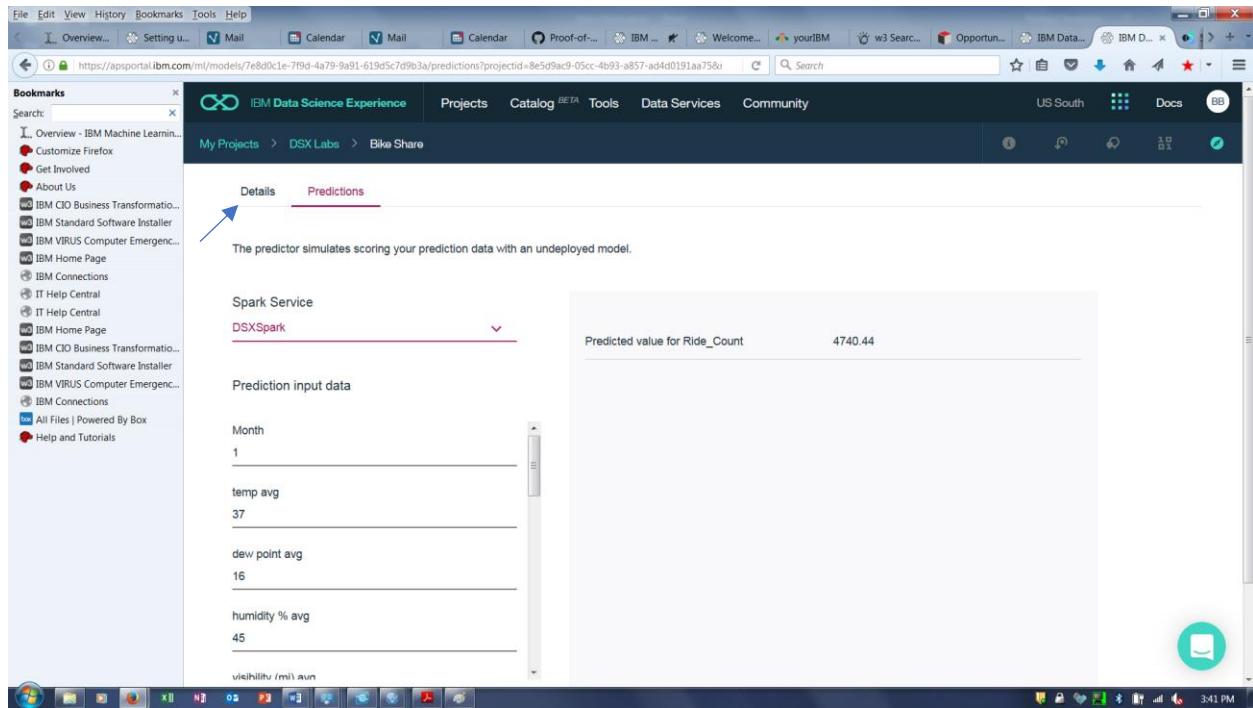
11. The estimated Ride_Count is then displayed

This screenshot is identical to the previous one, showing the prediction input form. However, a blue arrow points from the text 'Predicted value for Ride_Count' to the number '4352.03' displayed in the main content area. This indicates that the prediction has been successfully submitted and the result is now visible.

Deploying a Model

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

1. Select the Details Tab



The screenshot shows a web browser window for the IBM Data Science Experience. The URL in the address bar is <https://apsportal.ibm.com/ml/models/7e8d0c1e-7f9d-4a79-9a91-619d5c7d9b3a/predictions?projectid=8e5d9ac9-05cc-4b93-a857-ad4d0191aa75&o>. The page title is "IBM Data Science Experience". The left sidebar has a "Bookmarks" section with various IBM links. The main content area shows "My Projects > DSX Lab > Bike Share". There are two tabs: "Details" (selected) and "Predictions". A blue arrow points to the "Details" tab. Below it, a message says "The predictor simulates scoring your prediction data with an undeployed model." Under "Spark Service", it says "DSXSpark". In the "Prediction input data" section, there are five fields: "Month" (1), "temp avg" (37), "dew point avg" (16), "humidity % avg" (45), and "visibility /mi/ avg". To the right, a box shows "Predicted value for Ride_Count" as 4740.44. The bottom of the screen shows the Windows taskbar with various icons.

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

The screenshot shows the IBM Data Science Experience interface. On the left, there's a sidebar with bookmarks and a search bar. The main area displays the 'Bike Share' project under 'DSX Labs'. The 'Details' tab is selected, showing the following configuration:

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 this is a 'Deployments' section with a table:

NAME	DEPLOYMENT TYPE	ACTIONS
Your model is not deployed.		

An arrow points to the '+ Add Deployment' button at the top right of the table.

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

The screenshot shows the 'Deploy model' dialog box overlaid on the main interface. The dialog has the following fields:

- Deployment Type: Online (highlighted with a blue arrow)
- Name: Bike Share (highlighted with a blue arrow)
- Buttons: Close and Deploy (the Deploy button is highlighted with a red box and a blue arrow)

The background shows the same 'Bike Share' project details and deployment table 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 screenshot shows the IBM Data Science Experience web application. In the center, there's a modal window titled "Deployment Details" for a service named "Bike Share". The modal contains the following information:

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 this, under "Deployments", there's a table:

NAME	DEPLOYMENT TYPE	ACTIONS
Bike Share	Online	

At the top of the modal, a green checkmark icon and the message "Your deployment was successfully created." are displayed.

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

This screenshot shows the same IBM Data Science Experience interface, but the "Test API" tab is now selected in the navigation bar. The "Deployment Details" section remains the same as in the previous screenshot, displaying the deployment information for the "Bike Share" service.

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

The screenshot shows the IBM Data Science Experience interface. On the left, there's a sidebar with bookmarks and a search bar. The main area has tabs for 'Details' and 'Test API'. Under 'Test API', there's a section for 'Input data' with several input fields: 'Month' (1), 'temp avg' (37), 'dew point avg' (16), 'humidity % avg' (45), and 'visibility (mi) avg'. Below these is a 'Predict' button. A blue arrow points from the 'Predict' button towards the right side of the screen.

7. The predicted result is returned.

This screenshot shows the same interface as the previous one, but now the 'Predict' button has been clicked. In the bottom right corner of the input data section, the text 'Predicted value for Ride_Count' is displayed above the number '8541.35'. A blue arrow points from this predicted value back towards the center of the screen.

