

Overview

This lab will introduce the Watson Machine Learning capability using the Titanic 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 whether a person would survive
4. Deploying and Test 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:

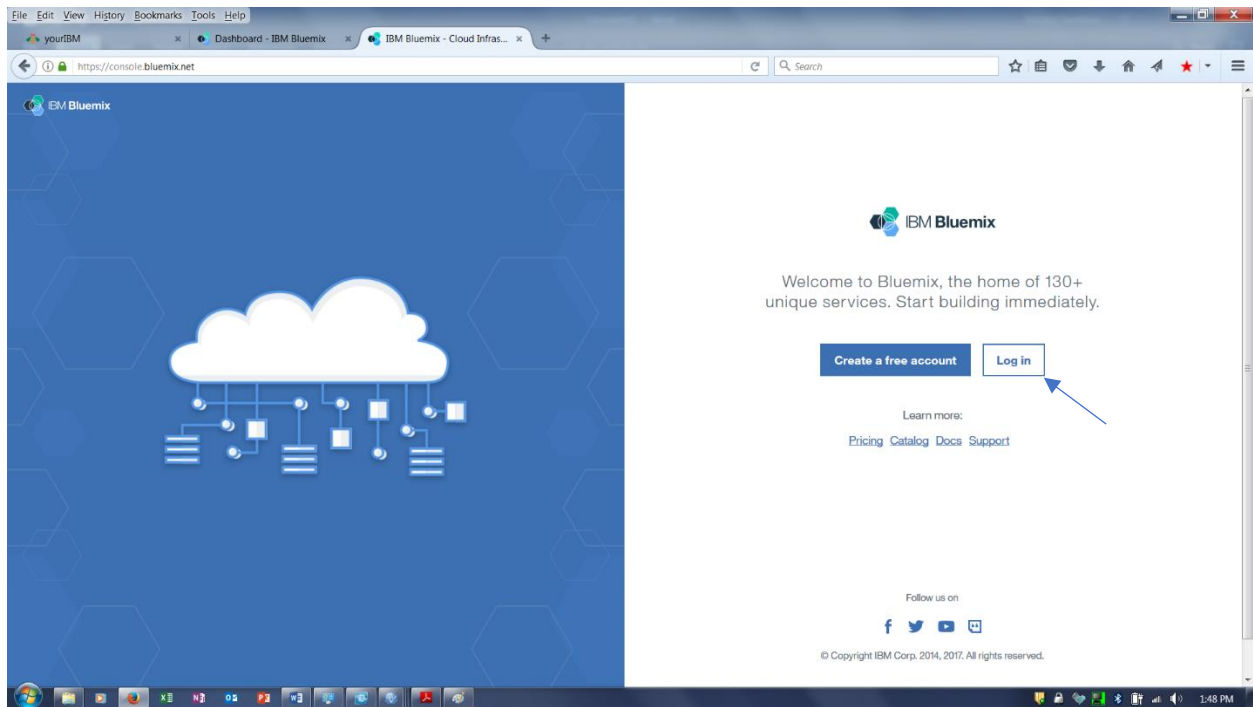
- Watson Machine Learning
- Object Storage
- Apache Spark

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.

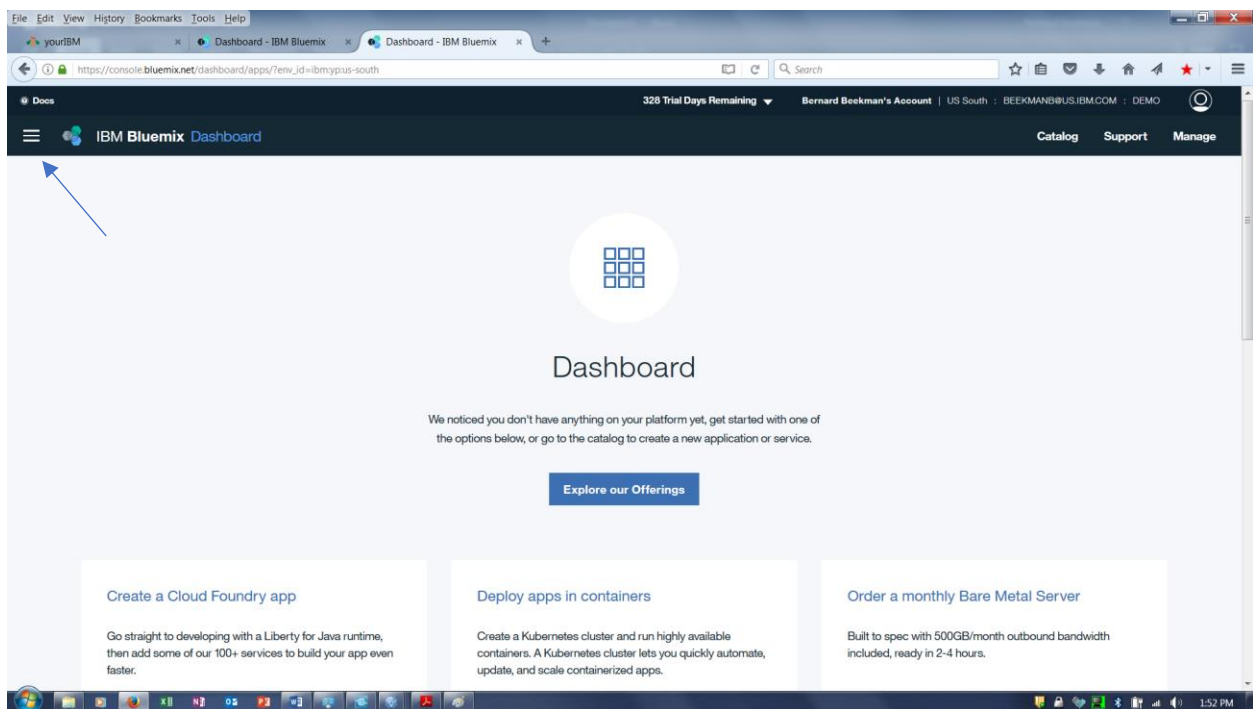
Step 1.1: Creating a Machine Learning Instance

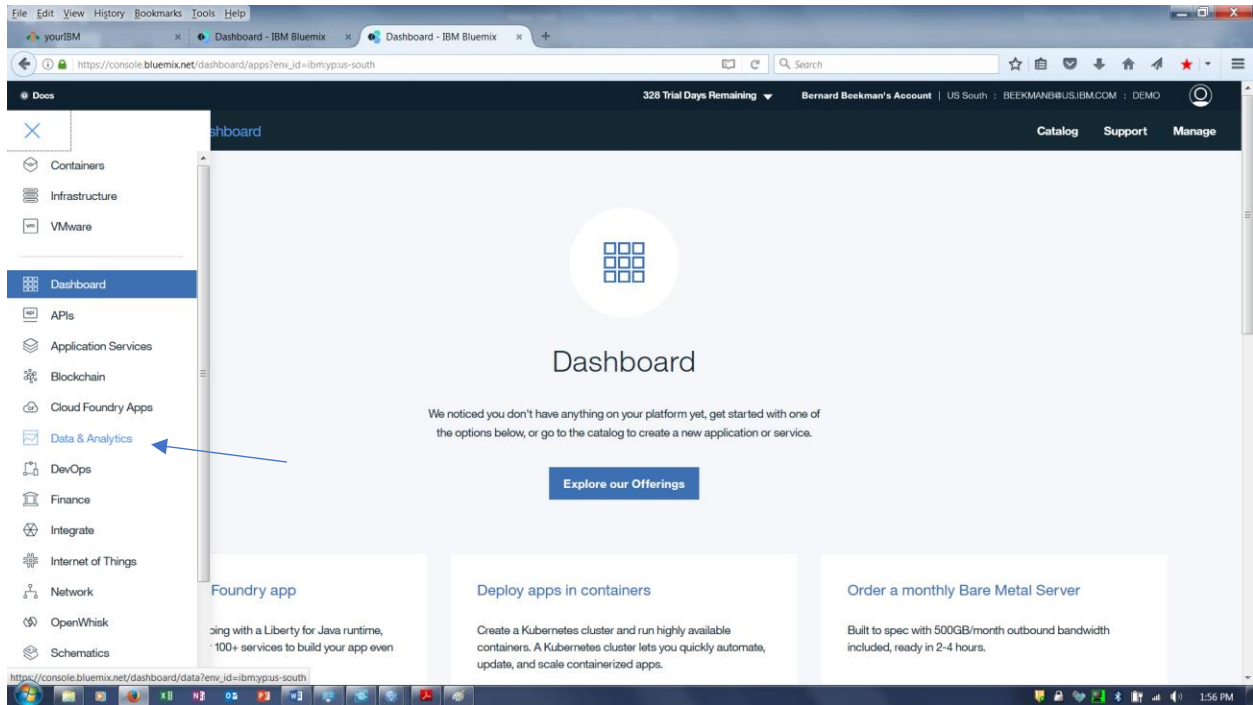
To create a Machine Learning service instance, you must perform the following steps:

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



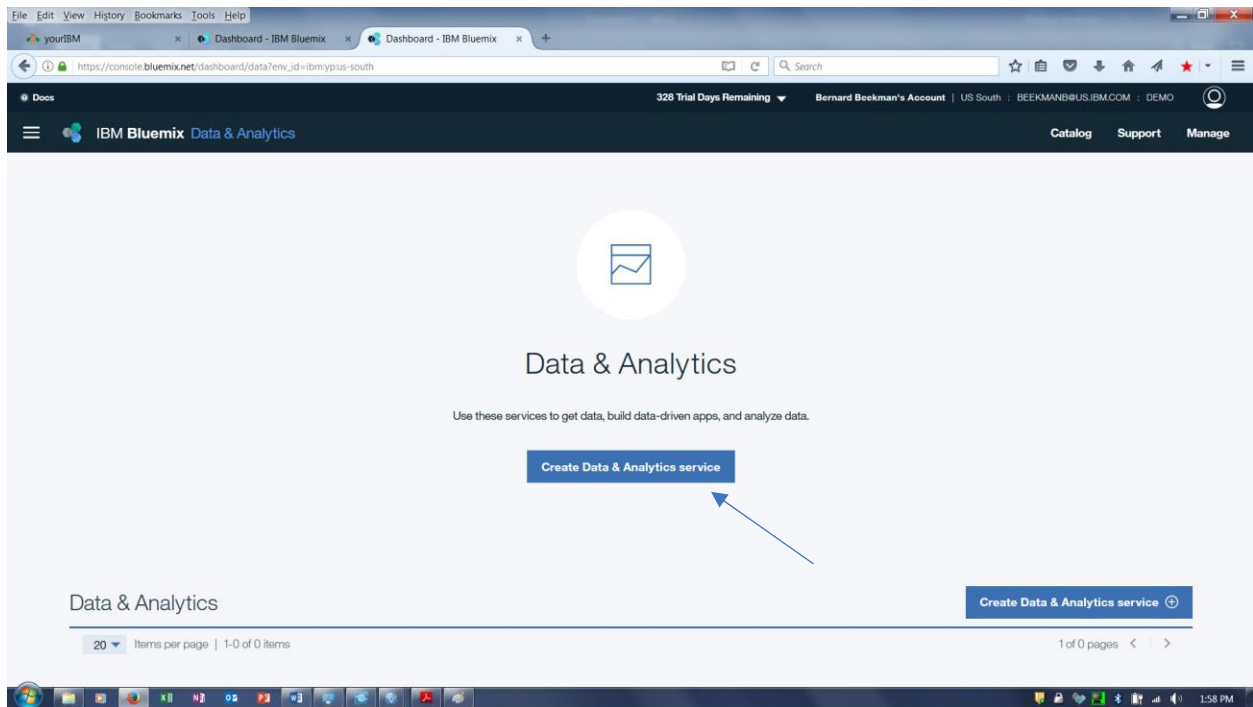
2. Once logged in, click on the hamburger icon, and from the navigation panel, click **Data & Analytics**.



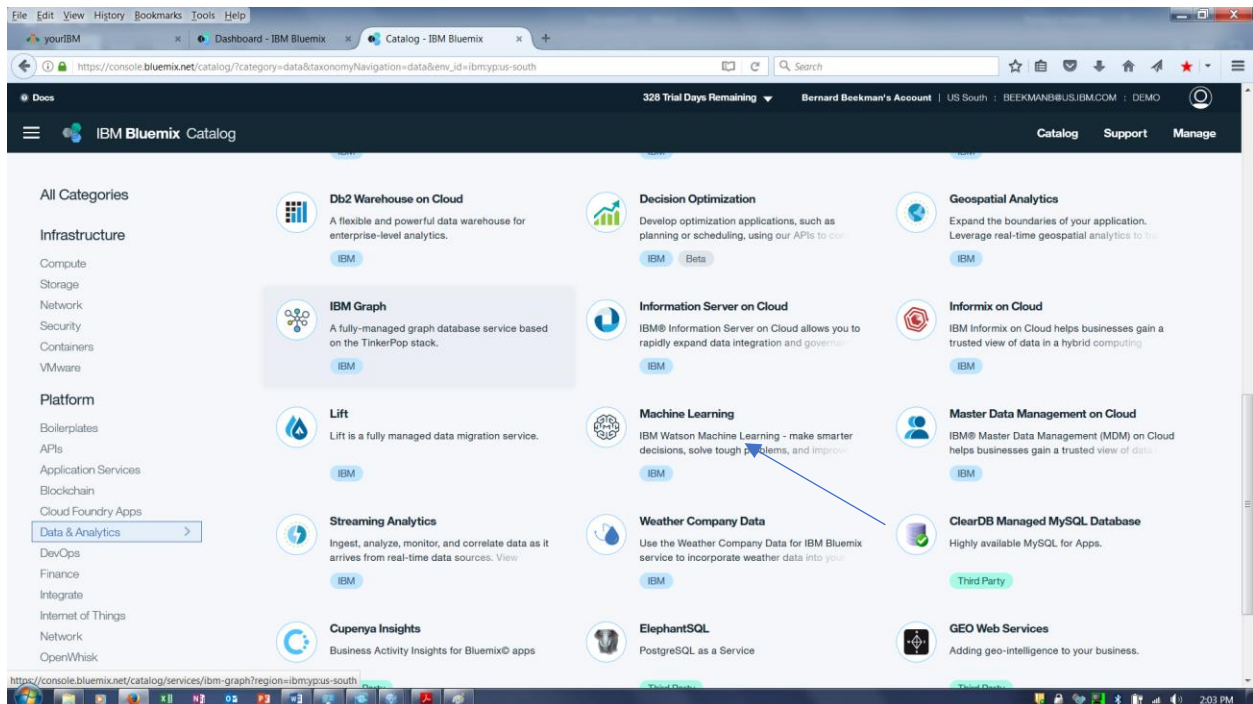


You see a screen centered on data services. You can return here regularly to work with your data and analytics services from one easy-to-use page. Check to see if a Machine Learning service already exists. If not, continue, otherwise go to Step 1.2: Adding existing Bluemix instances to a project in Data 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 description of the service, a metadata table, and a configuration form. Blue arrows highlight the following elements:

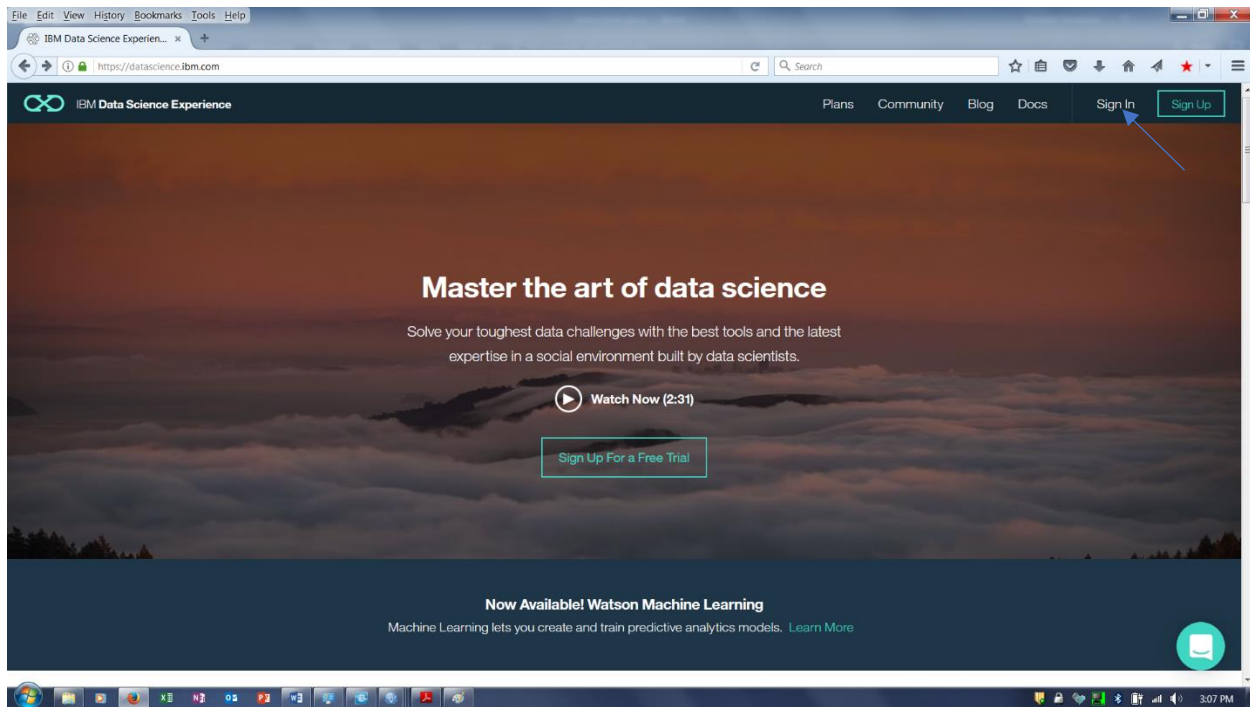
- Service name:** Machine Learning
- Credential name:** Credentials-1
- Select region to deploy in:** US South
- Choose an organization:** BEEKMANB@US.IBM.COM
- Choose a space:** DEMO
- Connect to:** Leave unbound
- Create button:** Located at the bottom right of the form.

Field	Value
AUTHOR	IBM
PUBLISHED	08/01/2017
TYPE	Service
LOCATION	US South, United Kingdom

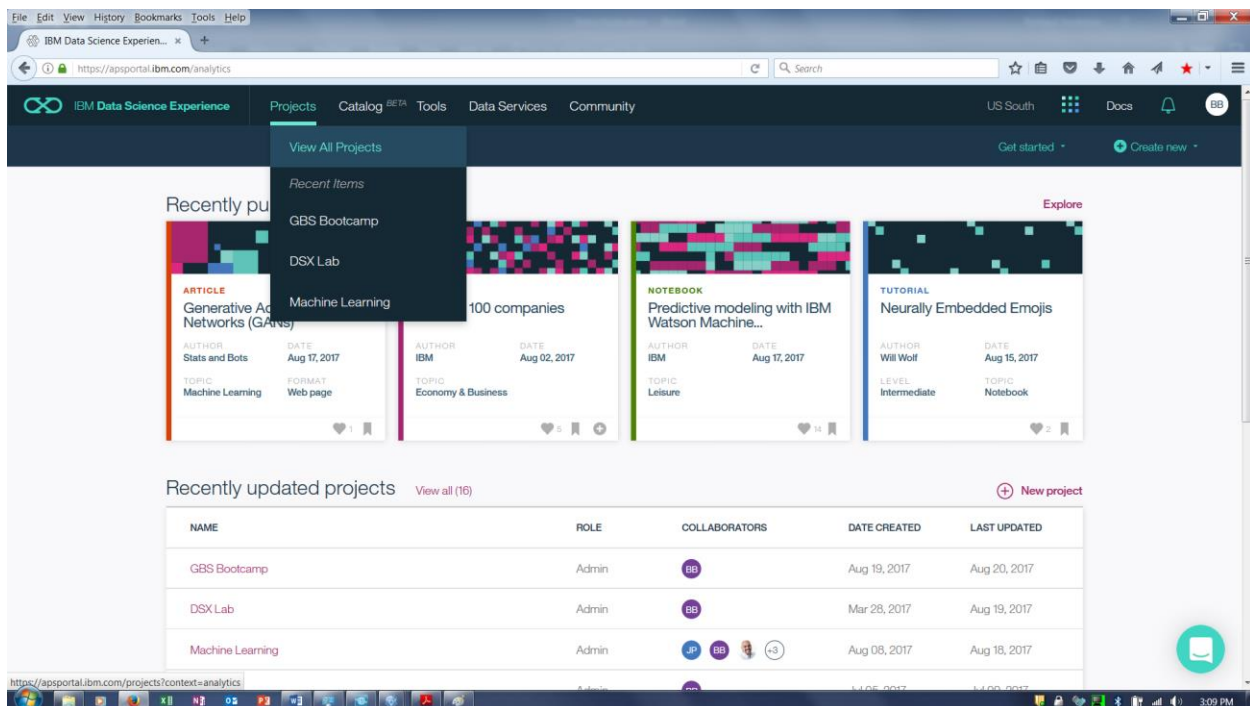
Step 1.2: Adding existing Bluemix instances to a project in Data Science Experience

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

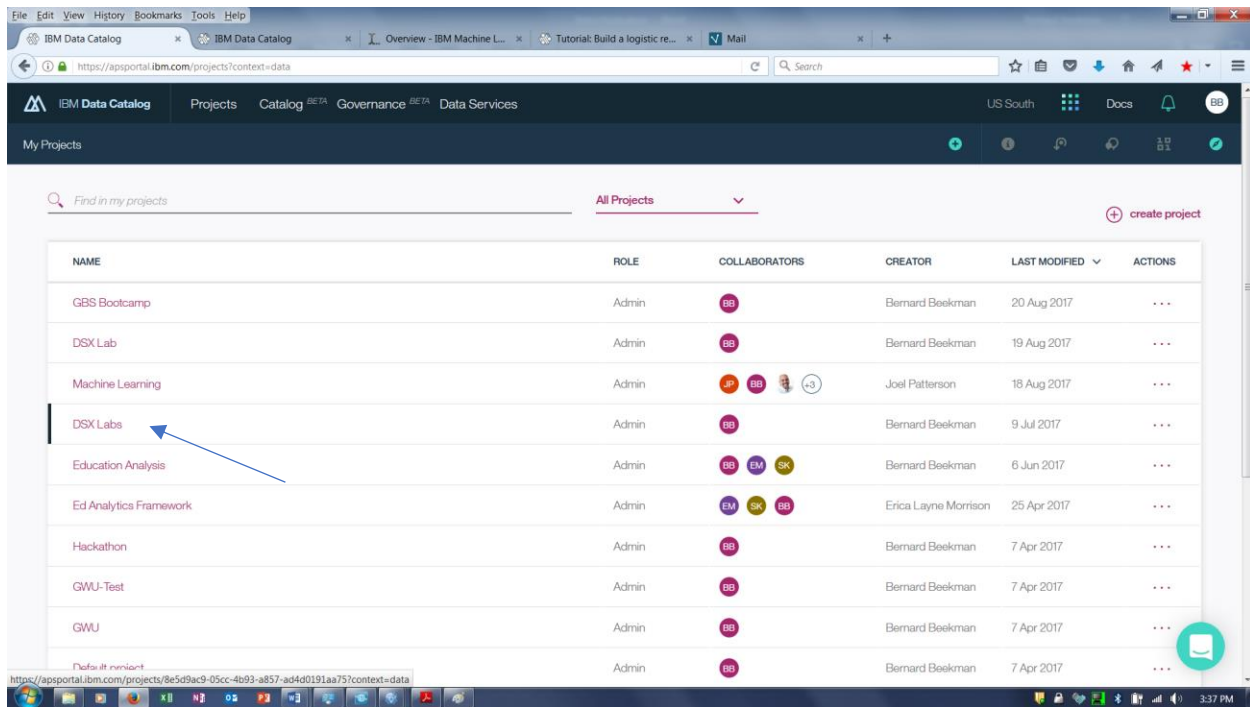
1. Log on to IBM Data Science Experience – <https://datascience.ibm.com>



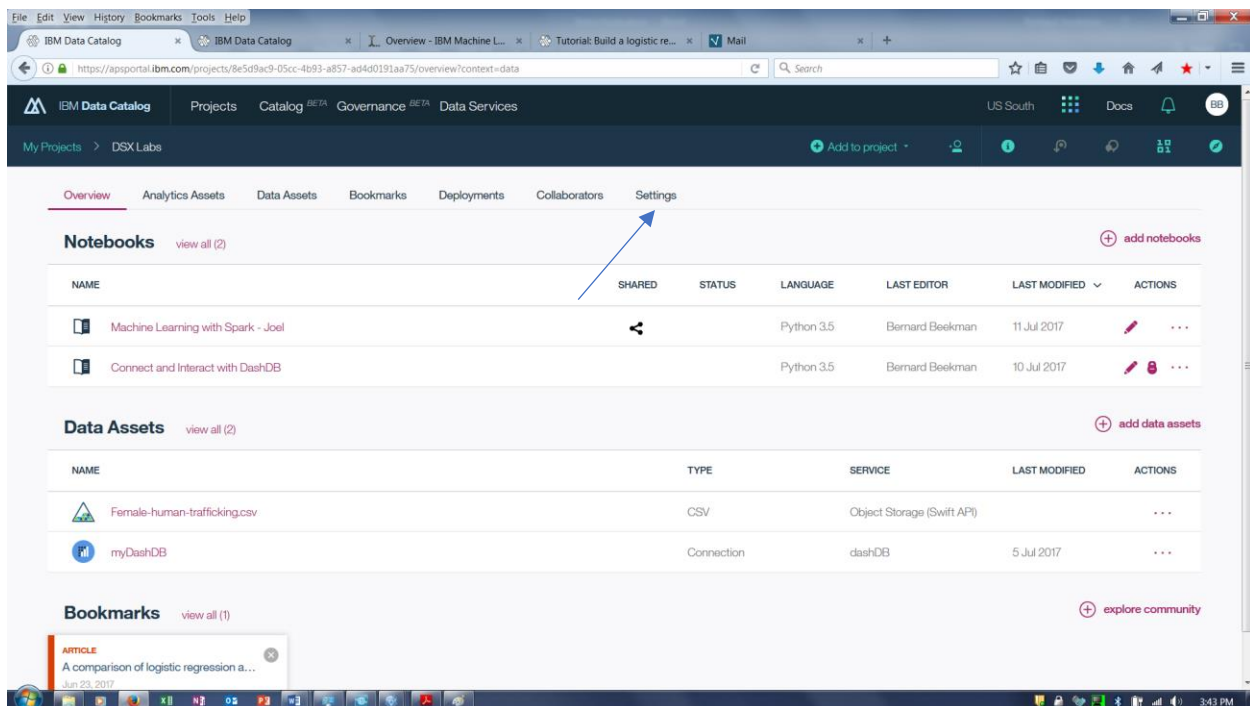
2. Click **Projects > View All Projects**.



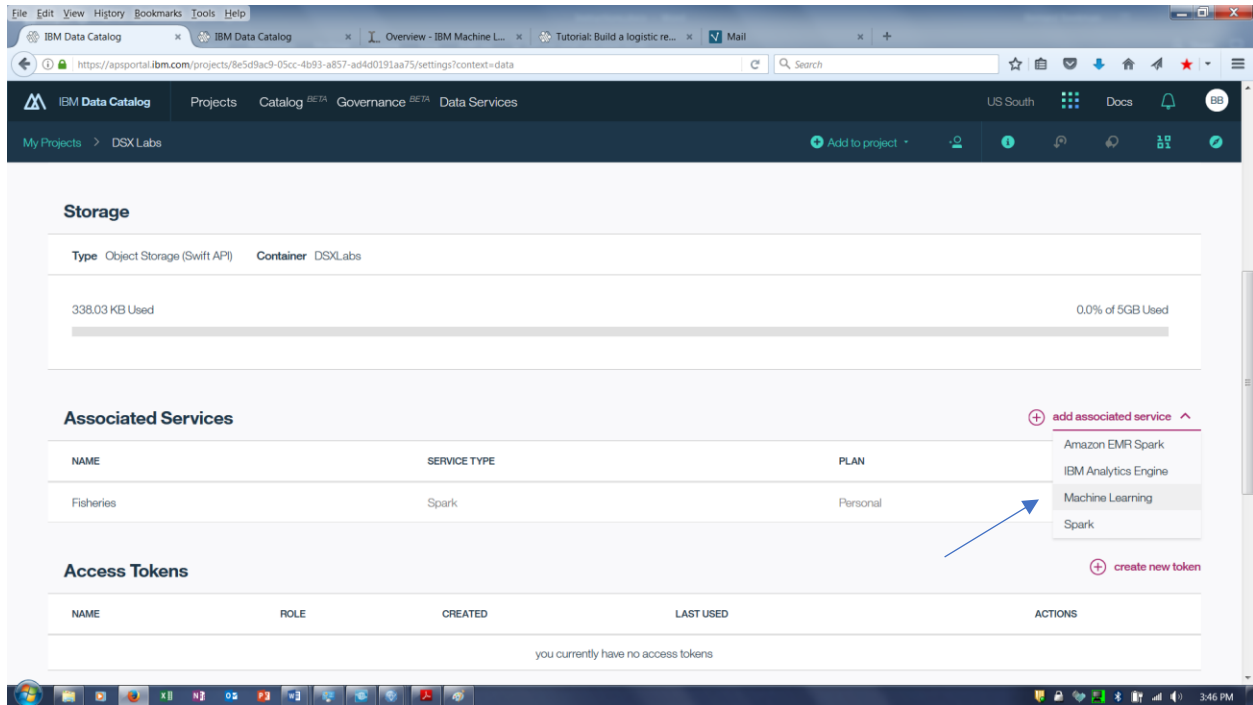
3. Click on the project that you created in the prerequisites, or if no project was created you can either create one, or click on the default project. (For the remainder of this document, I'm assuming the project name is DSX-Labs).



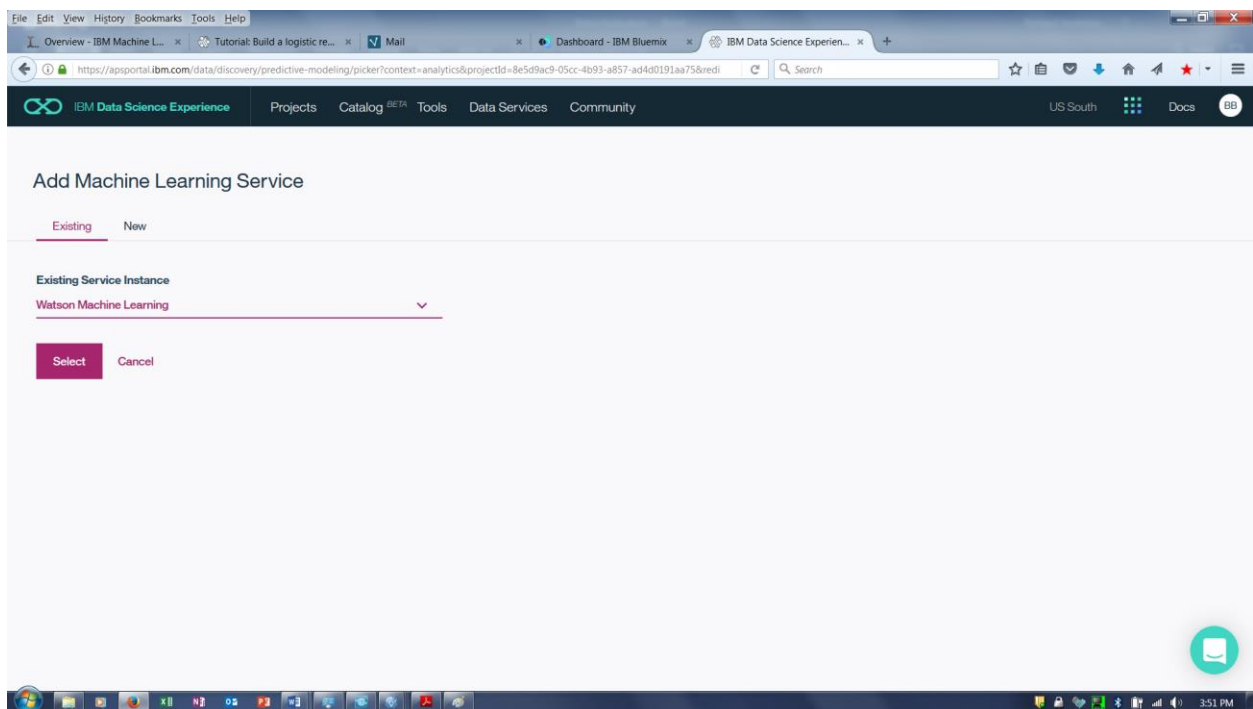
4. Select the **Settings** Tab.



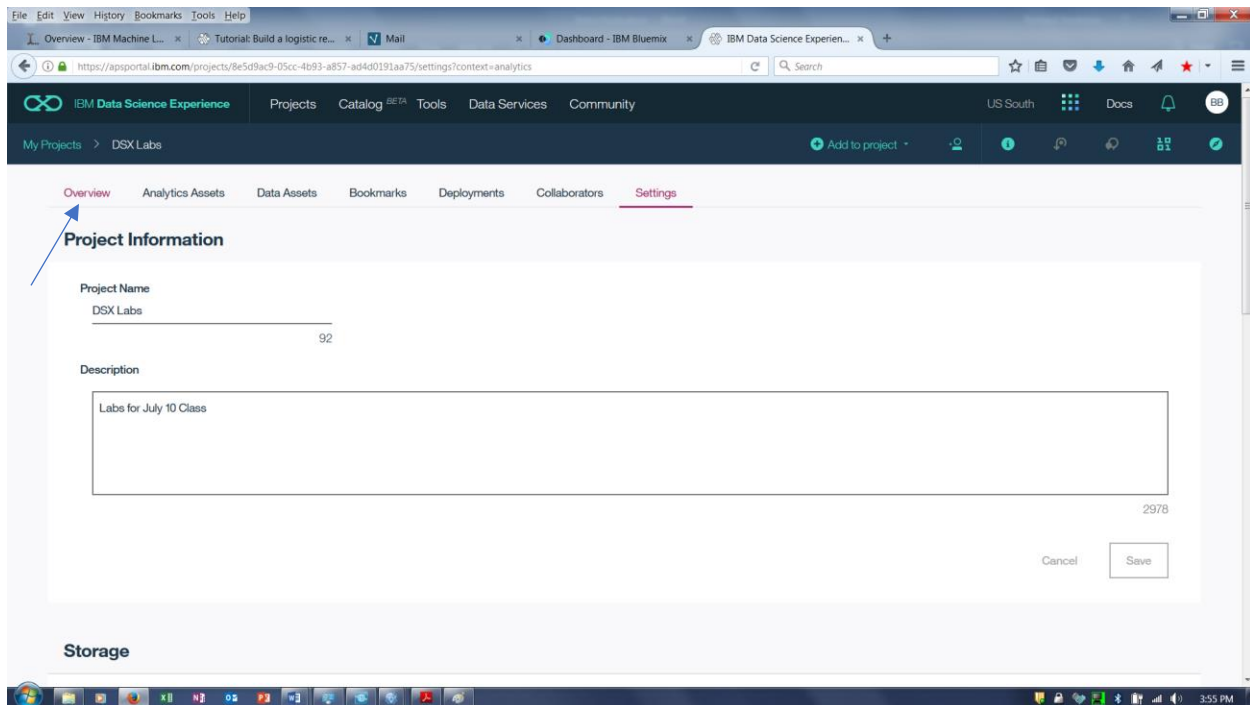
5. Scroll down to Associated Services. To add a service, in the **Associated Services** panel, click **add associated service**, select the Machine Learning service.



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



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



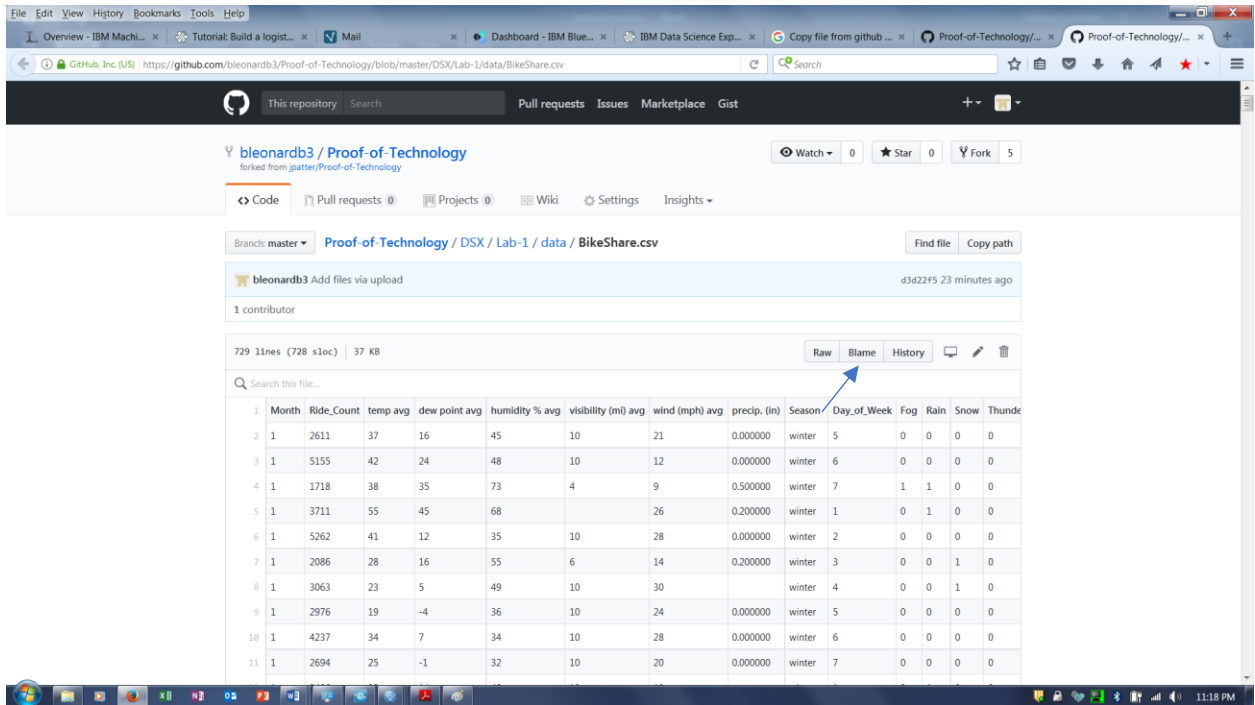
Step 2: Adding a Data Asset to the DSX Labs project

1. Download the Titanic data file from

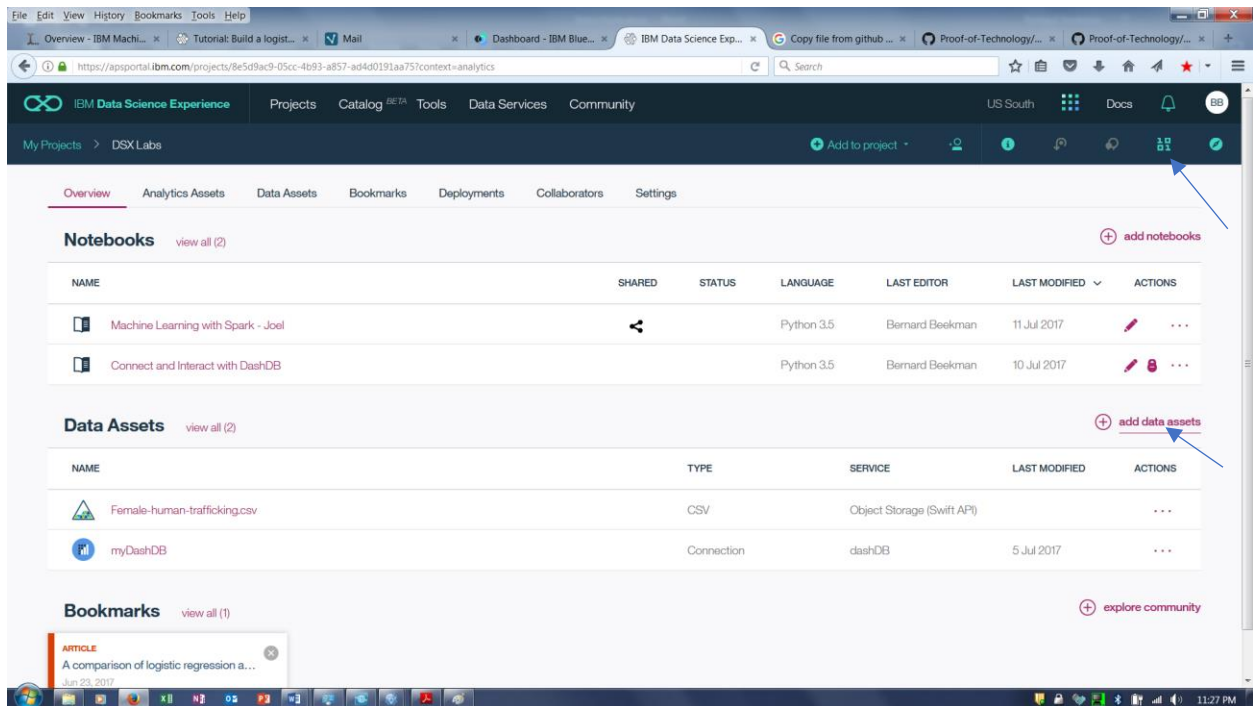
https://github.com/jpatter/ML-POT/data/titanic_cleansed.csv

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

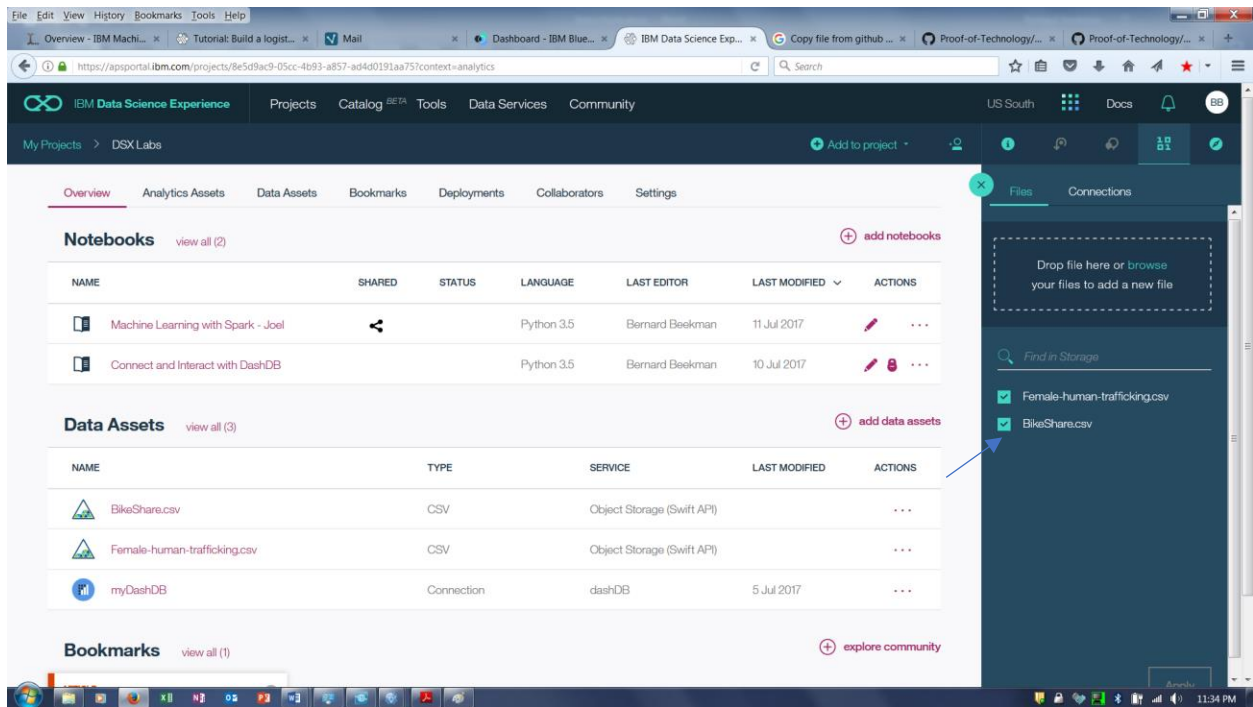
2. Right click on Raw, and click on Save link as



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

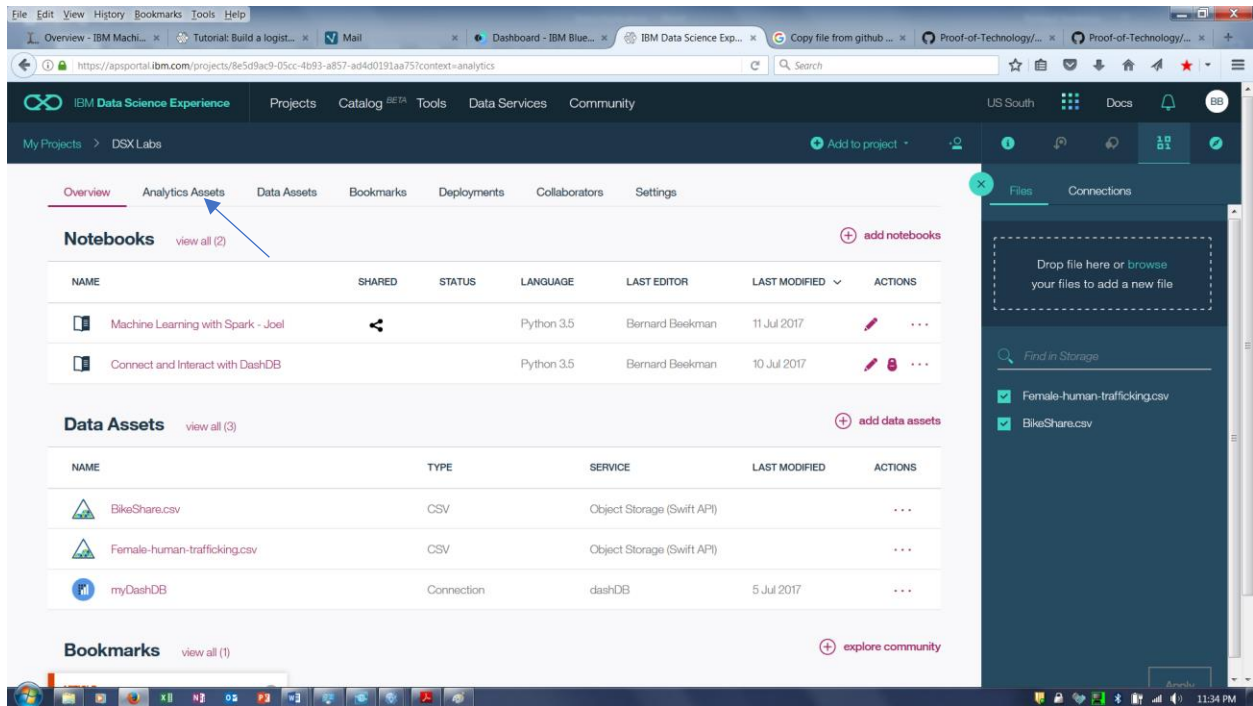


4. Click on browse and then go to the folder where the `titanic_cleansed.csv` is stored. Select `titanic_cleansed.csv` and then click Open.



Step 3: Create a Model to predict survival

1. Click on the Analytic Assets Tab



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

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 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 models listed: 'Female Human Trafficking- Manual' and 'Female Human Trafficking'. 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	...

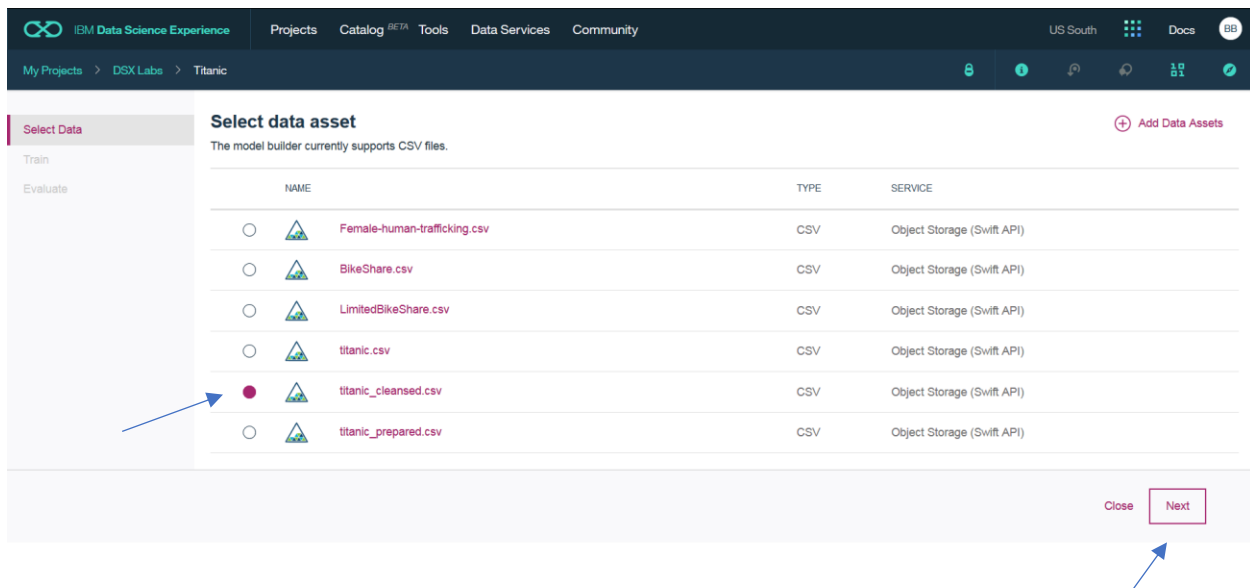
3. Enter the Model Name, Description, Select Manual, and click on **Create**.

The screenshot shows the 'Create new model' form. The form has the following fields and options:

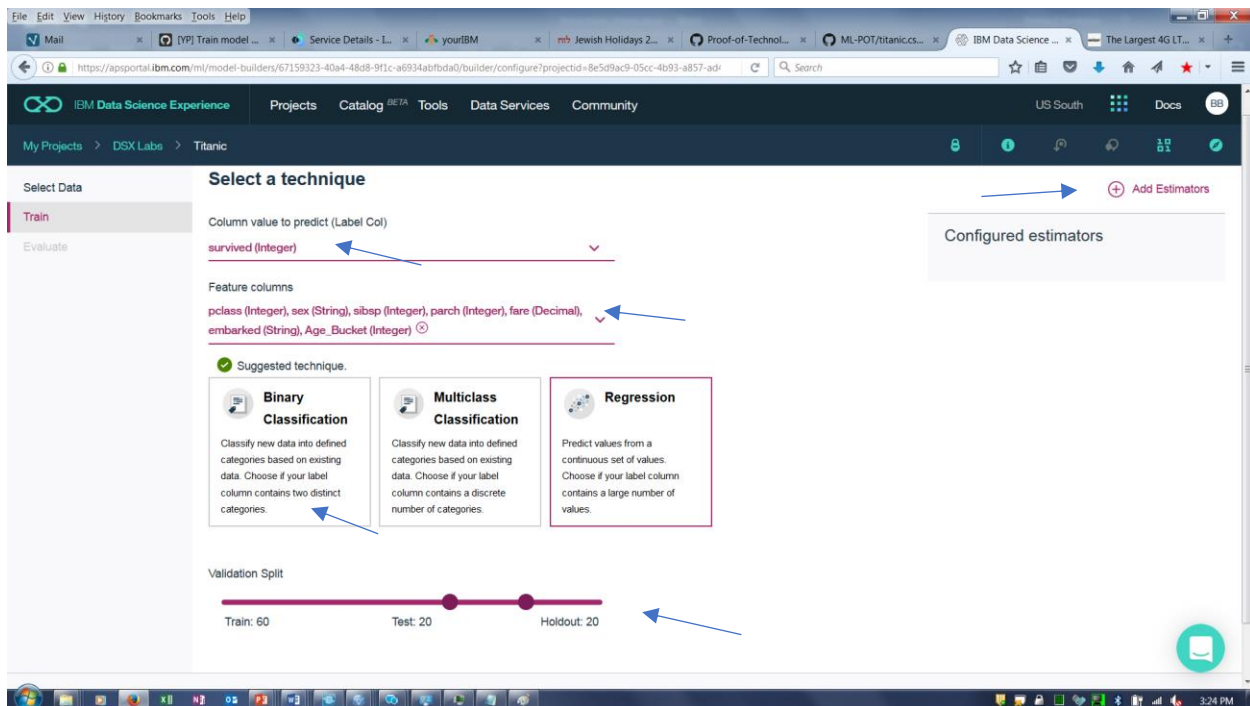
- Name:** Titanic
- Description:** Machine Learning Model for the Titanic Data Set
- Machine Learning Service:** Watson Machine Learning
- Spark Service:** DSXSpark
- Automatic:** Prepare my data and create a model automatically
- Manual:** Let me prepare my data and select which models to train

At the bottom right, there are 'Cancel' and 'Create' buttons. A blue arrow points to the 'Create' button.

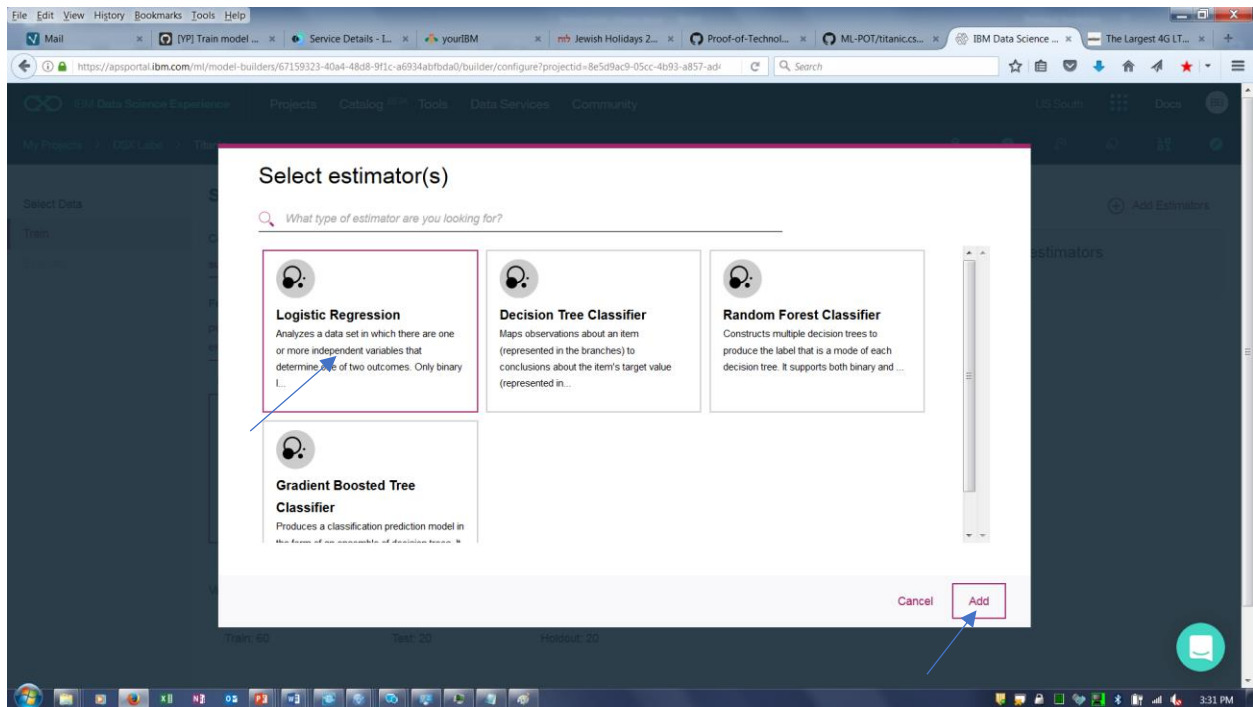
4. Click on the **titanic_cleansed.csv** and click on **Next**



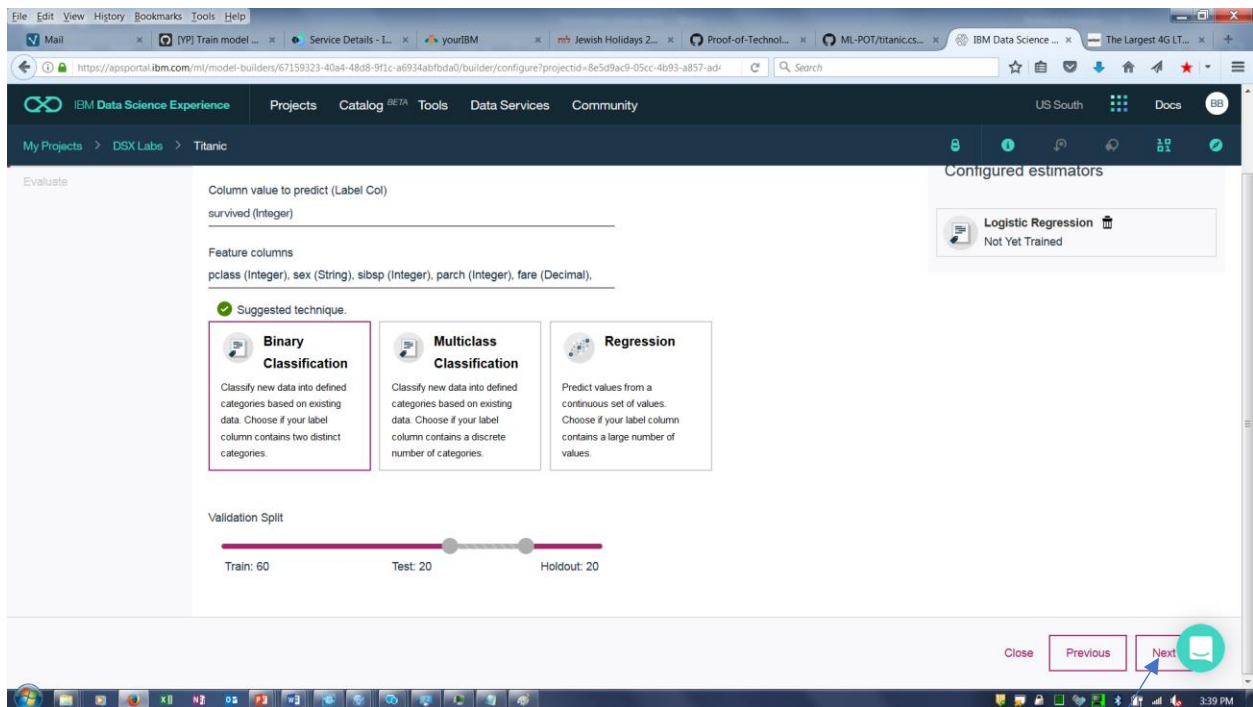
- For **Column value to predict (Label Col)** select **survivor**. For **Feature columns** select the following features (**pclass,sex,sibsp,parch,fare,embarked,Age_Bucket**) . Click on the **Binary Classification Box** (which is suggested by the service). Adjust the **Validation Split** as desired. Click on **Add Estimators** to add the specific models to use.



6. Select the **Logistic Regression**. Select **Add**.



7. Select the **Next** button.



8. The system trains and evaluates each model. If more than one model was selected, the models would be listed in ascending order of quality with the best result at the bottom. Click on the **Logistic Regression** and then click **Save**.

The screenshot shows the IBM Data Science Experience interface. The top navigation bar includes 'My Projects', 'DSX Labs', and 'Titanic'. The left sidebar has 'Select Data', 'Train', and 'Evaluate' tabs. The main area is titled 'Select model' and contains a table with the following data:

ESTIMATOR TYPE	STATUS	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST EVALUATION	ACTIONS
LogisticRegression	Trained & Evaluated	Good	0.81287	0.81673	26 Aug 2017, 6:29 PM	...

At the bottom right, there are buttons for 'Close', 'Previous', and 'Save'. A blue arrow points to the 'LogisticRegression' model in the table, and another blue arrow points to the 'Save' button.

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 top navigation bar includes 'My Projects', 'DSX Labs', and 'Titanic'. The left sidebar has 'Details' and 'Predictions' tabs. The main area is titled 'Titanic' and contains a table with the following data:

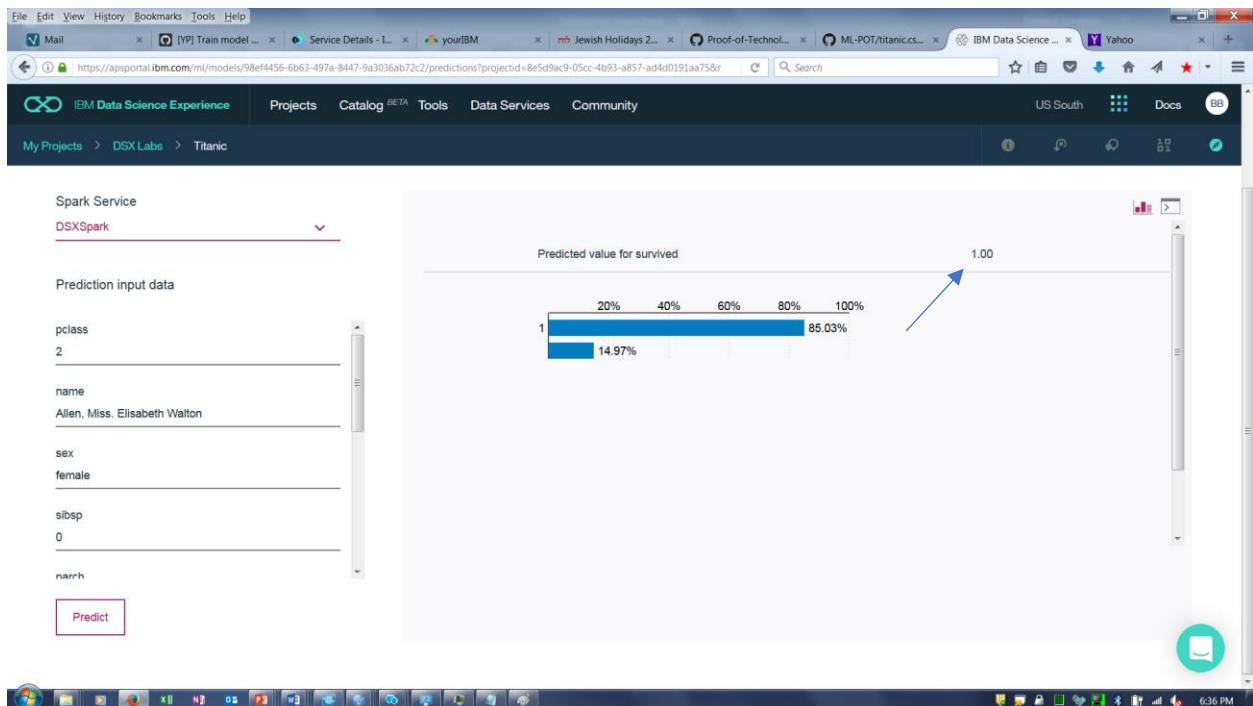
Machine learning service	Machine Learning-s2
Label column	survived
Model builder details	View
Training data schema	View
Input data schema	View
Runtime environment	spark-2.0
Training date	26 Aug 2017, 6:30 PM

Below the table, there is a section for 'Deployments'. A blue arrow points to the 'Predictions' tab in the sidebar.

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



11. The prediction for survivor is displayed along with the confidence in the prediction.



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 for a project named 'Titanic'. The top navigation bar includes 'Projects', 'Catalog', 'Tools', 'Data Services', and 'Community'. The left sidebar shows 'My Projects' > 'DSX Labs' > 'Titanic'. The main content area displays project details:

Label column	survived
Model builder details	View
Training data schema	View
Input data schema	View
Runtime environment	spark-2.0
Training date	26 Aug 2017, 6:30 PM

Below the details is the 'Deployments' section. It features a table with columns 'NAME', 'DEPLOYMENT TYPE', and 'ACTIONS'. The table is currently empty, with the message 'Your model is not deployed.' displayed below it. An 'Add Deployment' button is visible in the top right corner of the deployments section, indicated by a blue arrow.

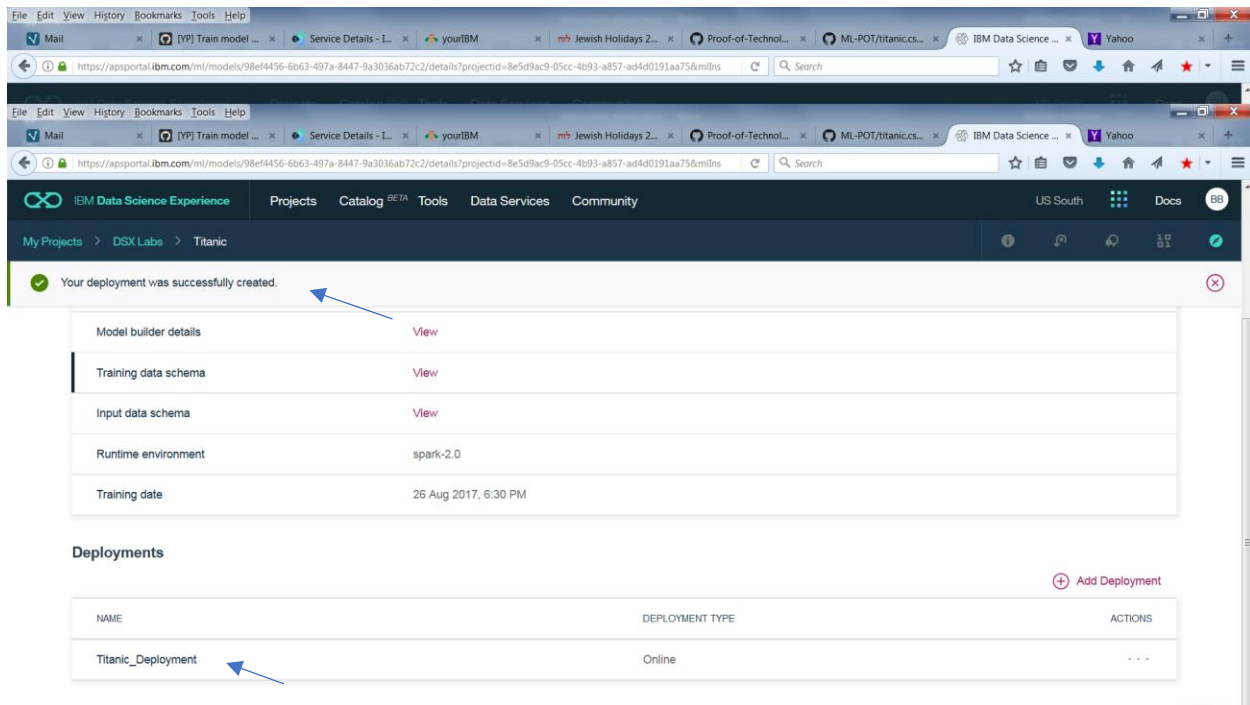
3. Select Online for **Deployment Type**, enter Titanic_Deployment for **Name**, and click on **Deploy**.

The screenshot shows the 'Deploy model' dialog box open in the IBM Data Science Experience interface. The dialog box has the following fields and buttons:

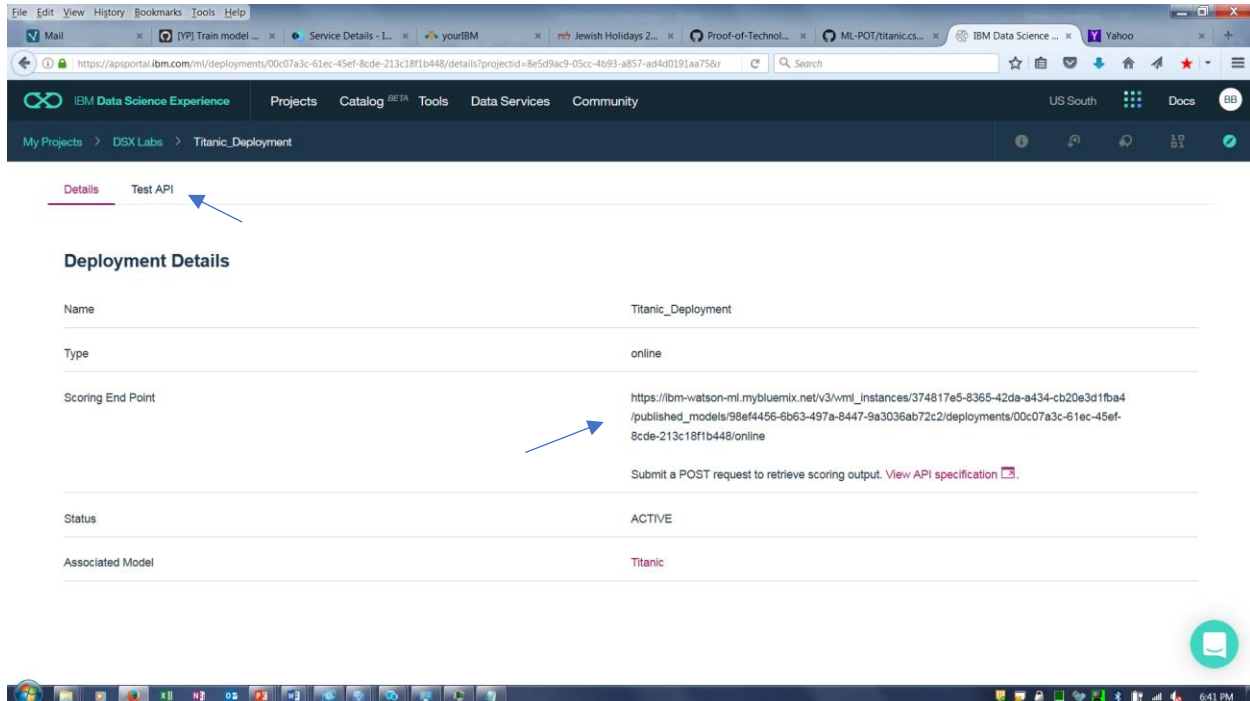
- Deployment Type:** A dropdown menu with 'Online' selected. A blue arrow points to this dropdown.
- Name:** A text input field containing 'Titanic_Deployment'. A blue arrow points to this field.
- Buttons:** 'Close' and 'Deploy' buttons are at the bottom right. A blue arrow points to the 'Deploy' button.

The background shows the same project details as the previous screenshot, but the 'Deployments' section is partially obscured by the dialog box.

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



- The system displays information about the deployed service including the endpoint to invoke by an application (e.g web application predicting survival). Click on **Test API** to test out the API.



- Enter values for the input fields and then click on **Predict**. Note that the values inputted for any of the fields not included in the model parameters (e.g. name) will not affect the prediction.

My Projects > DSX Labs > Titanic_Deployment

Details Test API

Input data

pclass
2

name
Allen, Miss. Elisabeth Walton

sex
female

sibsp
0

narch

Predict



7. The predicted result is returned.

