Operating AI with Trust and Transparency

Michael Cronk, IBM, michael.cronk@ibm.com

Bernard Beekman, IBM, beekmanb@us.ibm.com

Table of Contents

Disclaimer
IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results like those stated here.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts.
In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal

advice or represent or warrant that its services or products will ensure that the customer follows any law.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

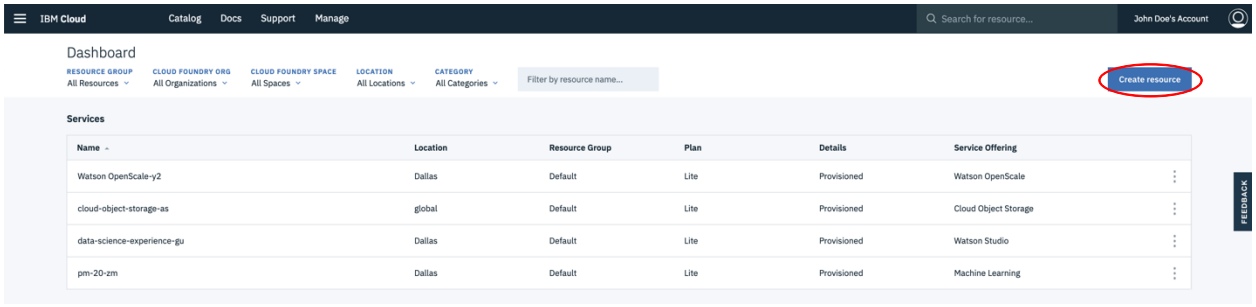
The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, ibm.com and [names of other referenced IBM products and services used in the presentation] are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

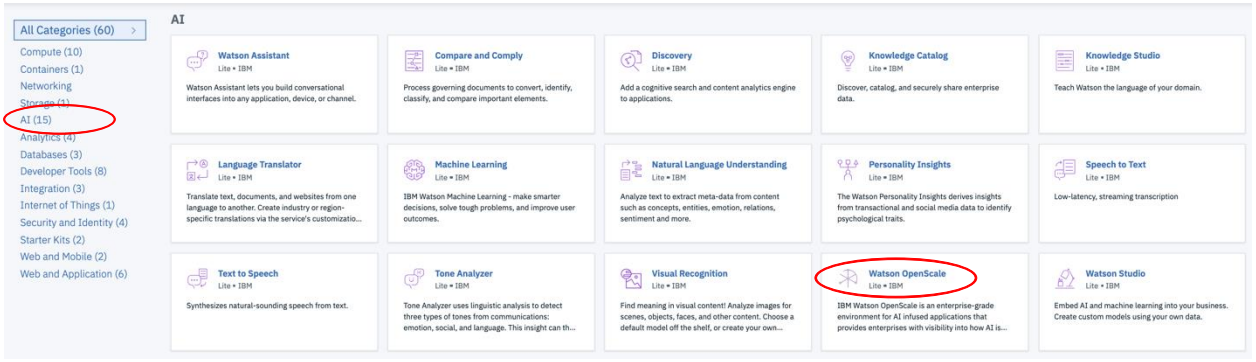
U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

Step 1: Create a Watson OpenScale service.

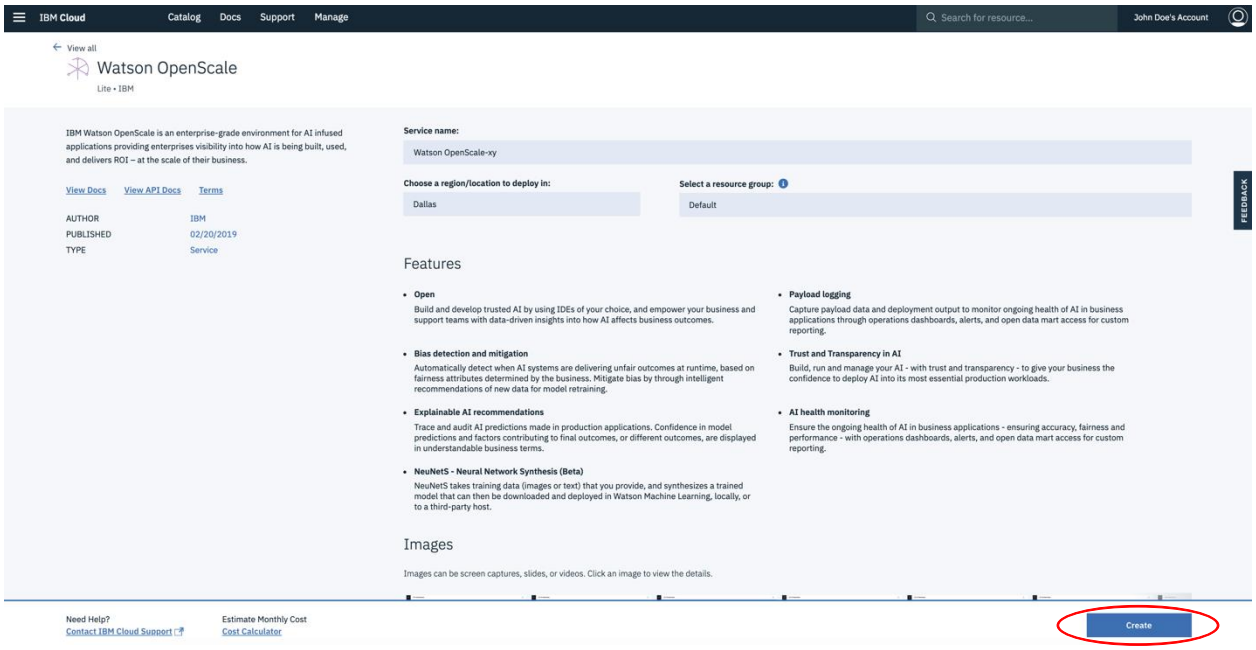
1. Open up a new tab and access bluemix.net
2. Click on Create resource.



3. Along the left side, click on AI.
4. Scroll down (if necessary) and click on Watson OpenScale.



5. Once the Watson OpenScale page opens, keep the defaults and click Create.

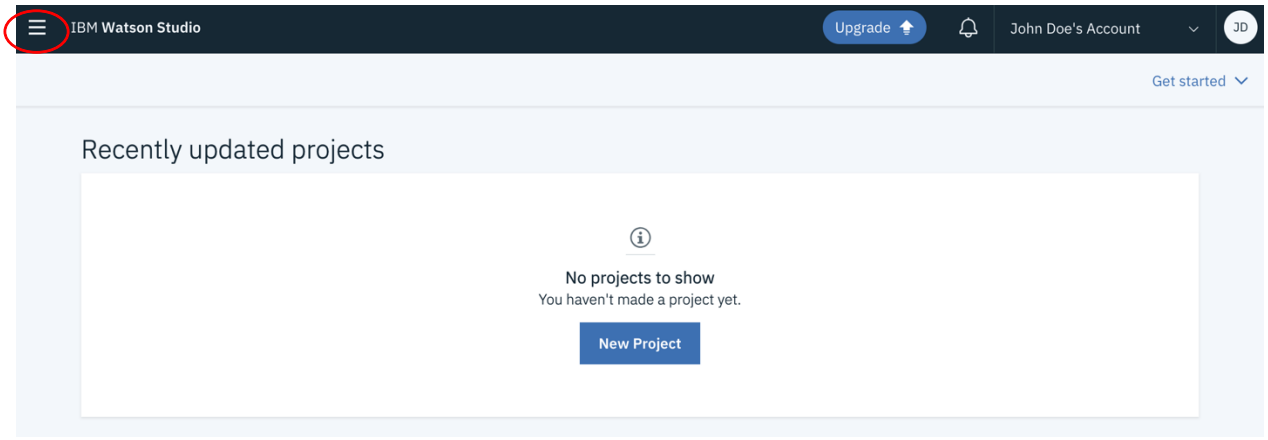


Step 2: Create a Watson Studio Project.

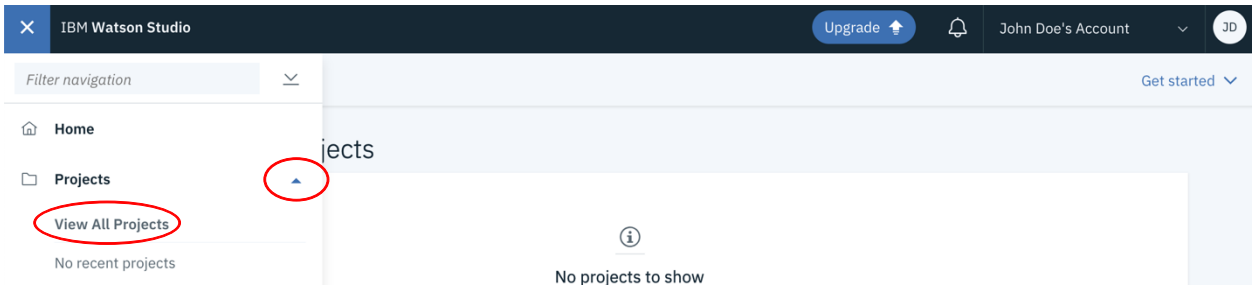
Watson OpenScale connects to a deployed machine learning model. We will use Watson Studio (a data science workbench) and Watson Machine Learning (a machine learning deployment engine). Note, Watson Machine Learning is not required to create and deploy a machine learning model for Watson OpenScale. OpenScale also supports Azure ML, AWS Sagemaker, as well as custom models.

Before we begin, we need to create a Watson Studio Project. A Watson Studio Project is a collection of assets geared towards a specific data science goal. These assets can include machine learning models, runtimes, data files, and connections to 3rd party/open source databases.

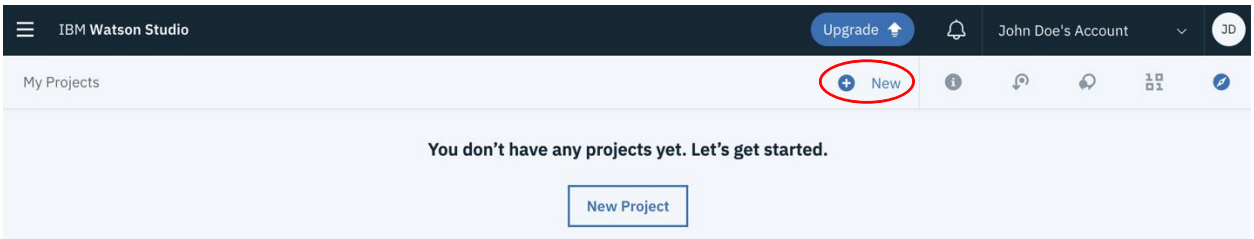
1. Access dataplatform.cloud.ibm.com.
2. Click on the Hamburger menu located in the top left of the screen.



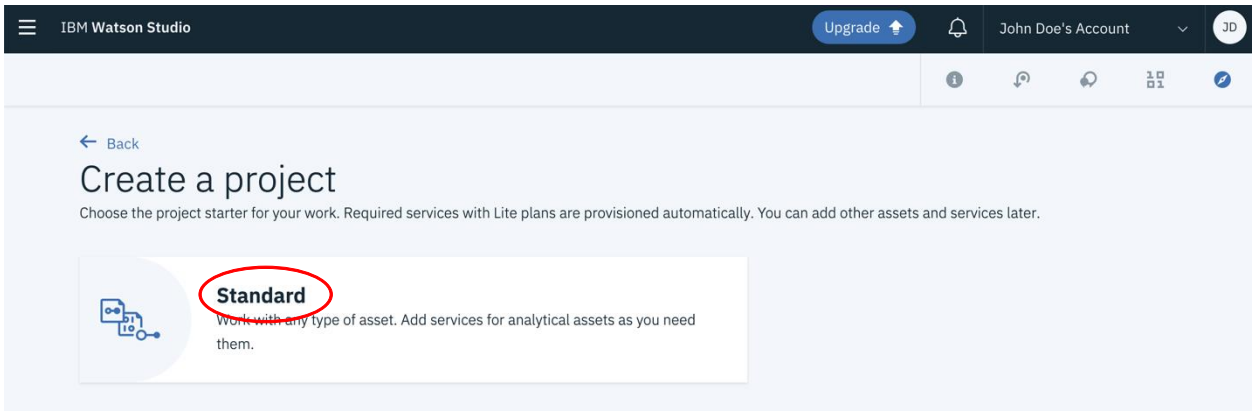
3. Click on Projects -> View all Projects.



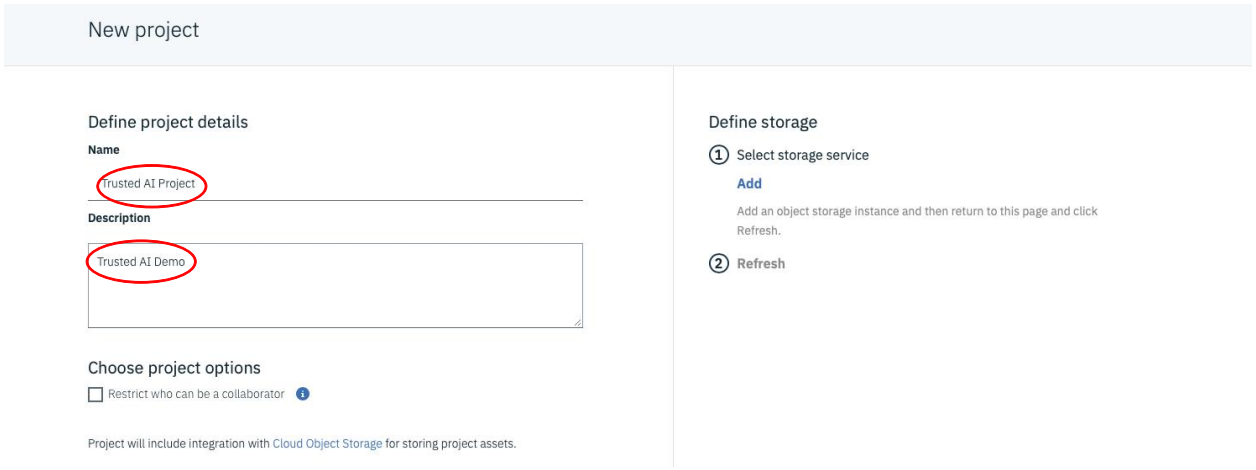
4. Click on New.



5. Click on Standard.

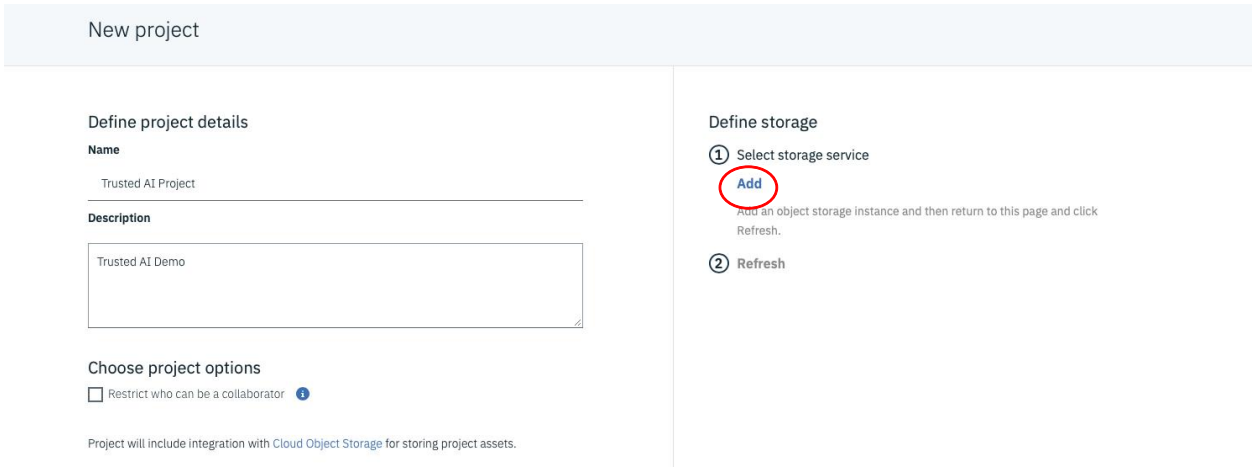


6. Give your project a Name, a Description (optional), and click Create.

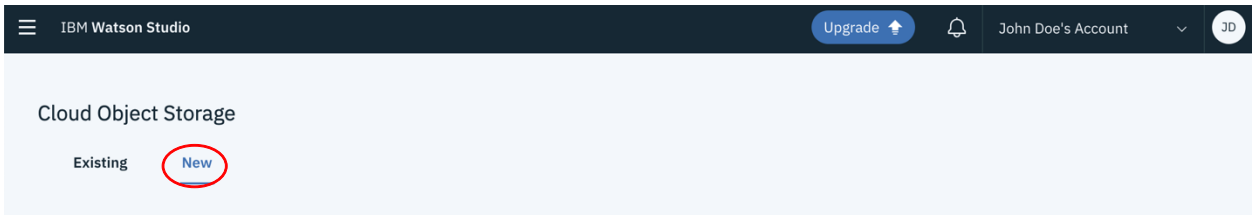


Cloud object storage is the default store for basic project assets. We will requisition a small instance for our project.

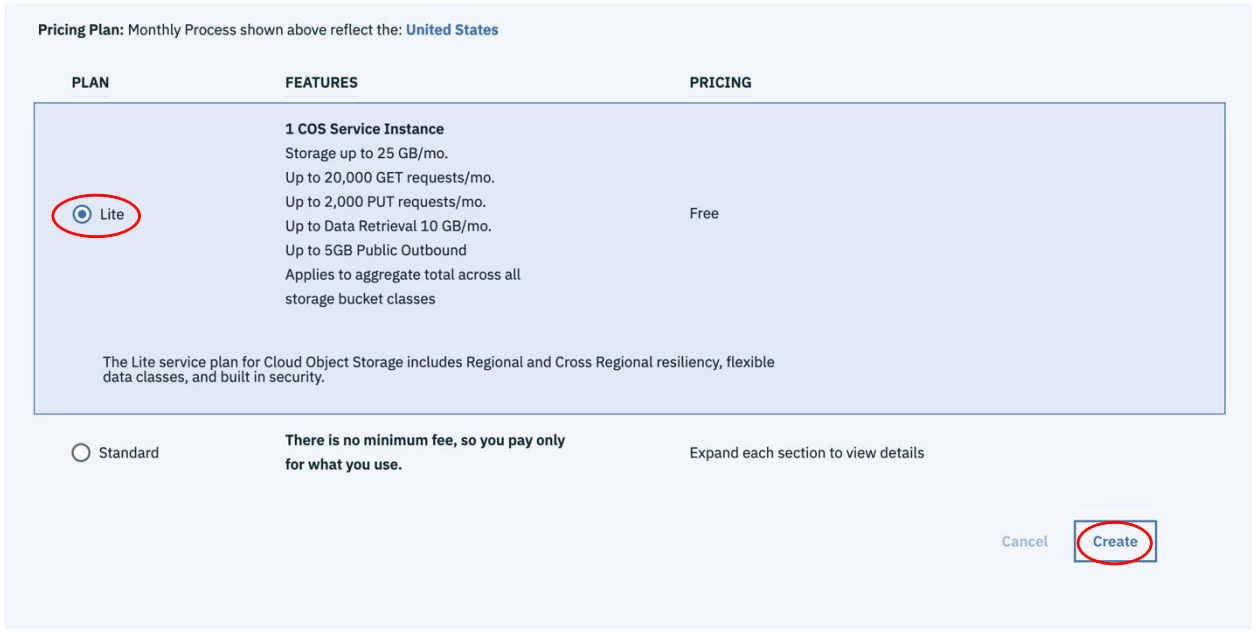
7. On the right side of the screen, click on Add under Define storage.



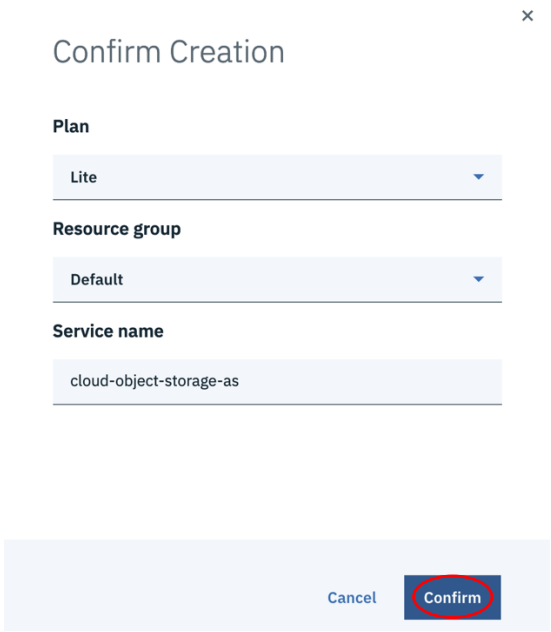
8. Click on New.



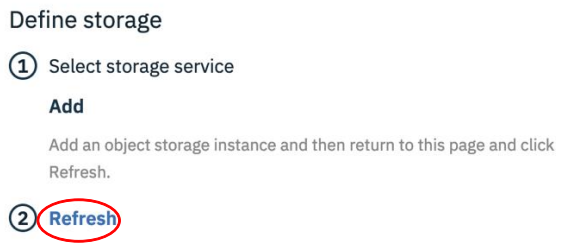
9. Scroll down, click on Lite, and click Create.



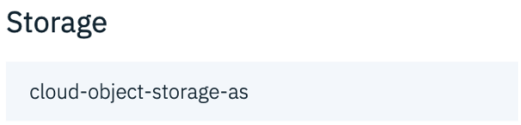
10. Click on Confirm.



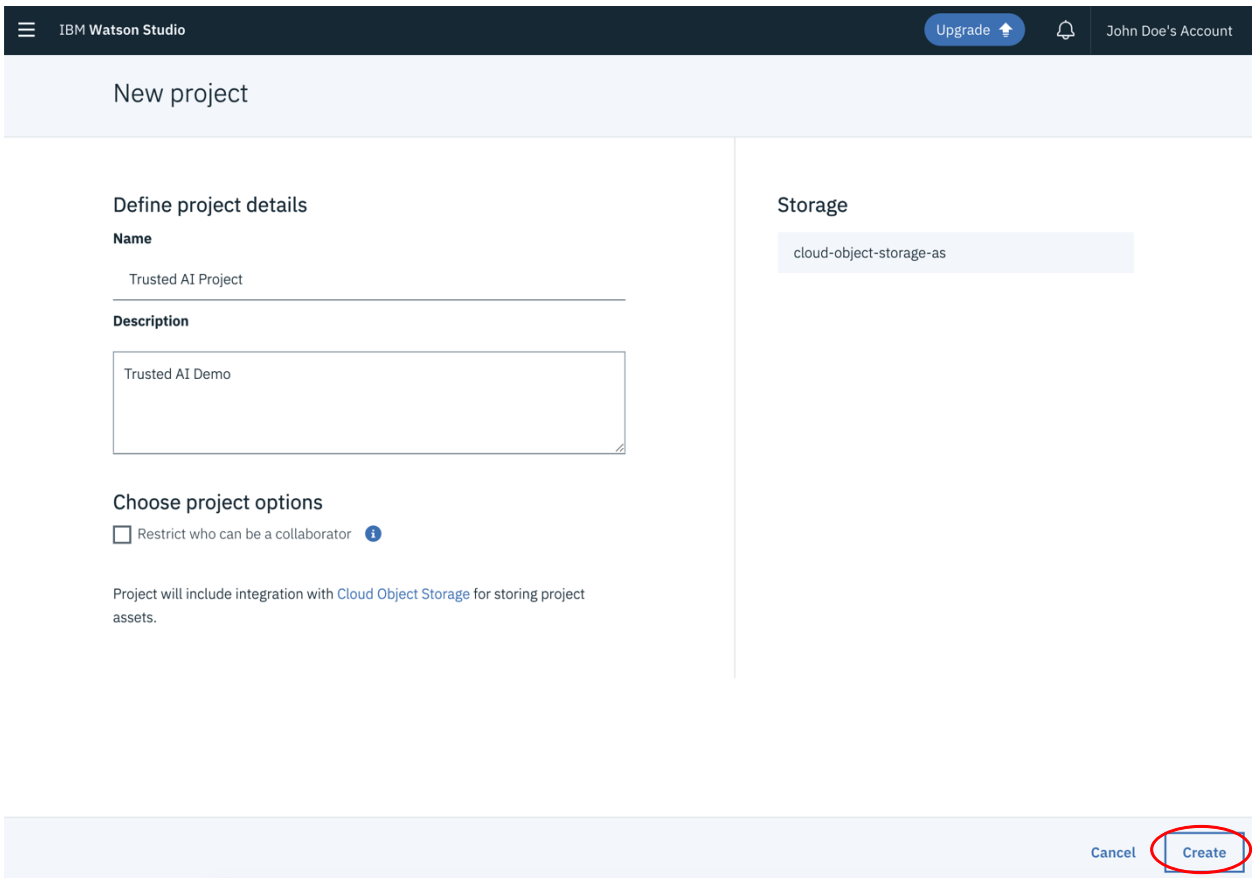
11. Once you are brought back to the New project page, click Refresh.



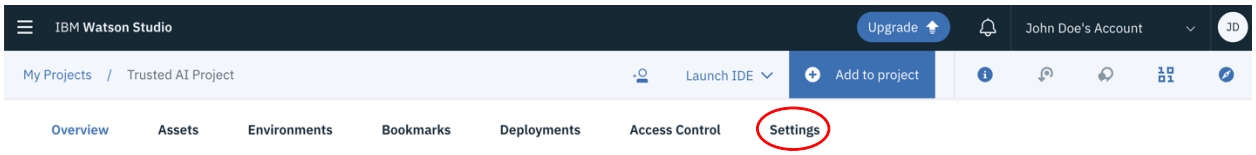
The name of the storage service you have just created should appear under Storage.



12. Click Create.

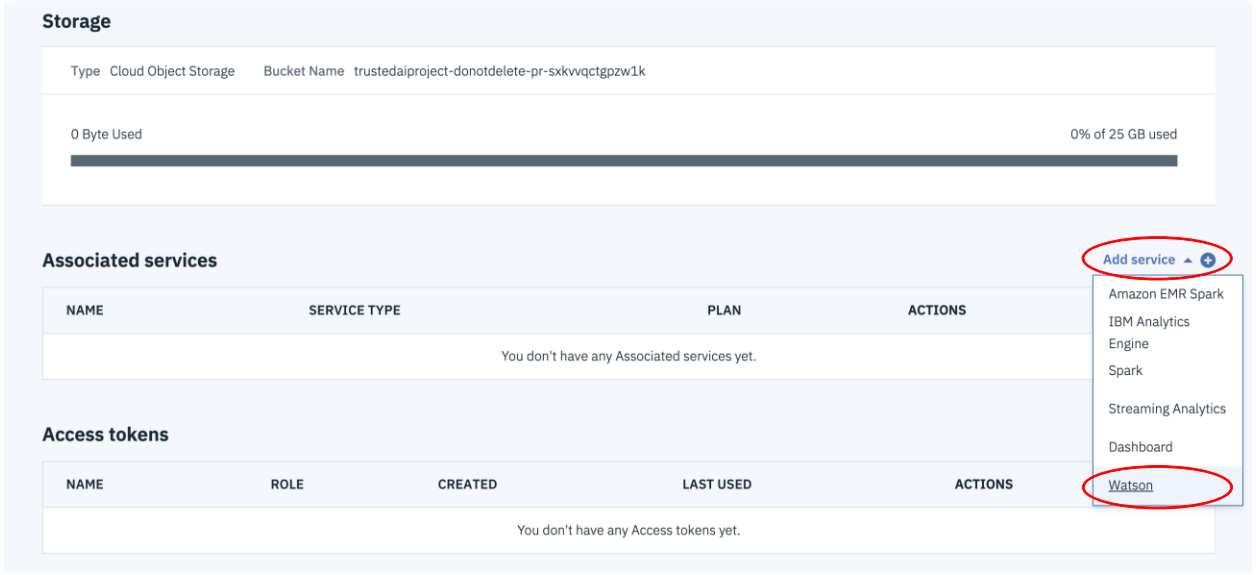


13. Click on the Settings tab.

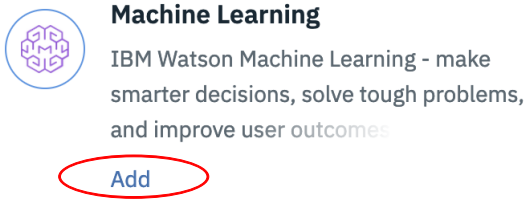


14. Scroll down and click on Add a service for Associated services.

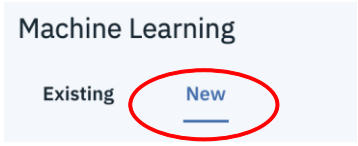
15. Click on Watson.



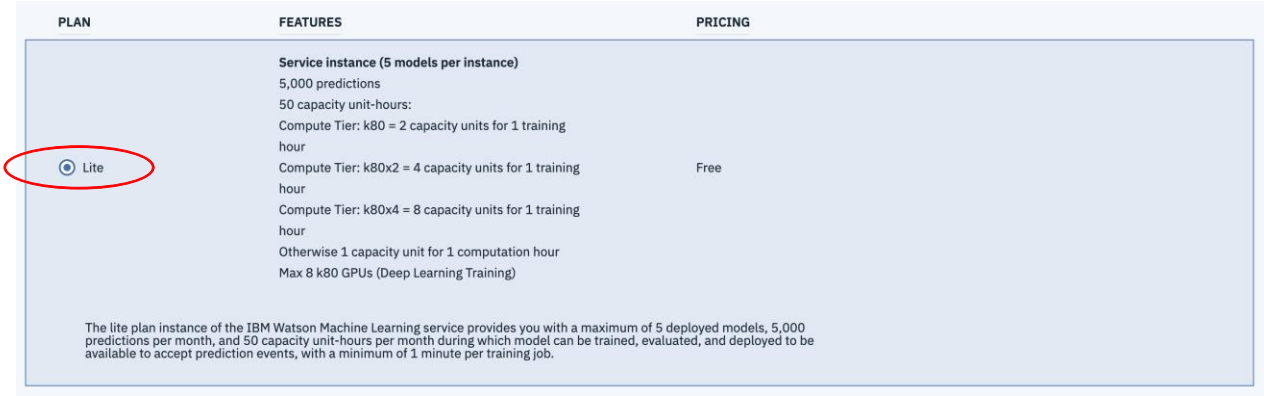
16. Click on Add under Machine Learning.



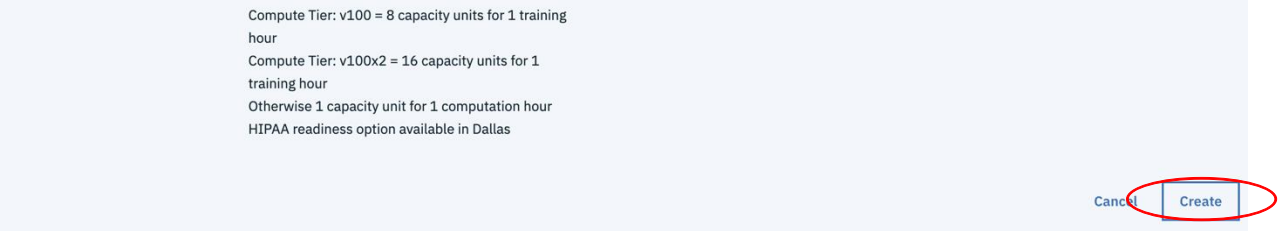
17. Ensure that the New tab is selected.



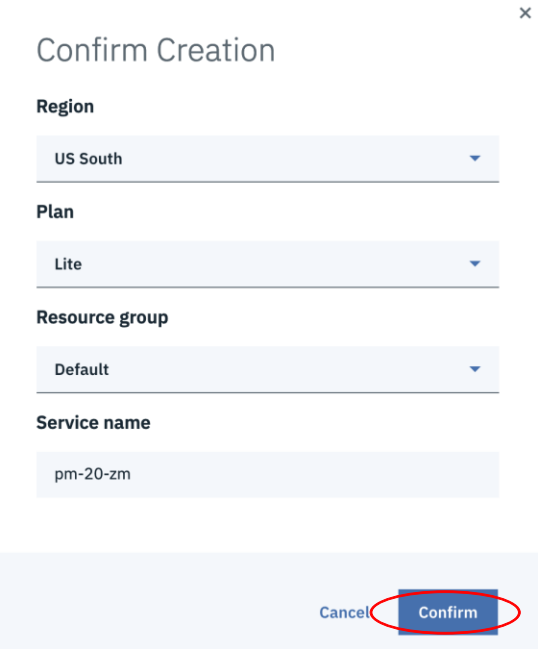
18. Scroll down and click on Lite.



19. Click on Create.

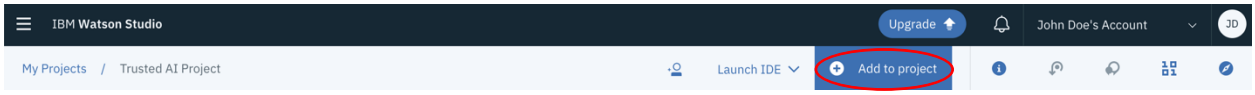


20. Keep the defaults and Click Confirm.

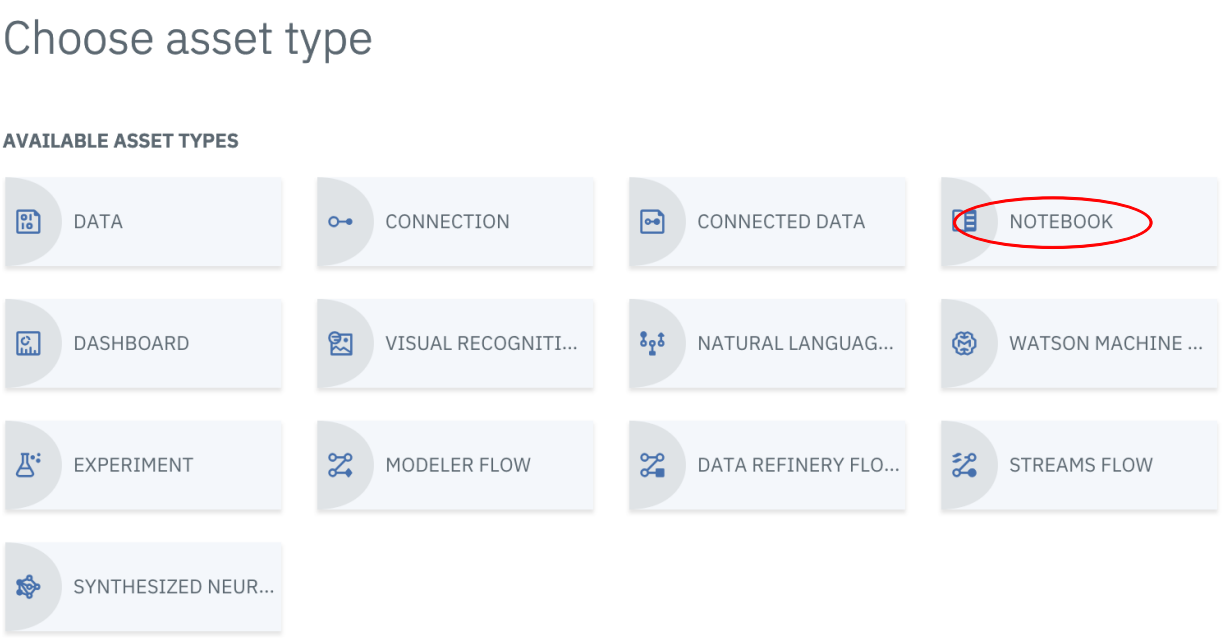


Step 3: Set up Jupyter Notebooks and Configure Watson OpenScale.
A Jupyter Notebook is a self-contained environment that combines documentation and code. This allows us to walk through code and document each step. We will use one Jupyter Notebook to build and deploy a machine learning model and then configure OpenScale to monitor that deployed model. A second notebook will be used to automatically feed data to OpenScale to monitor for bias.

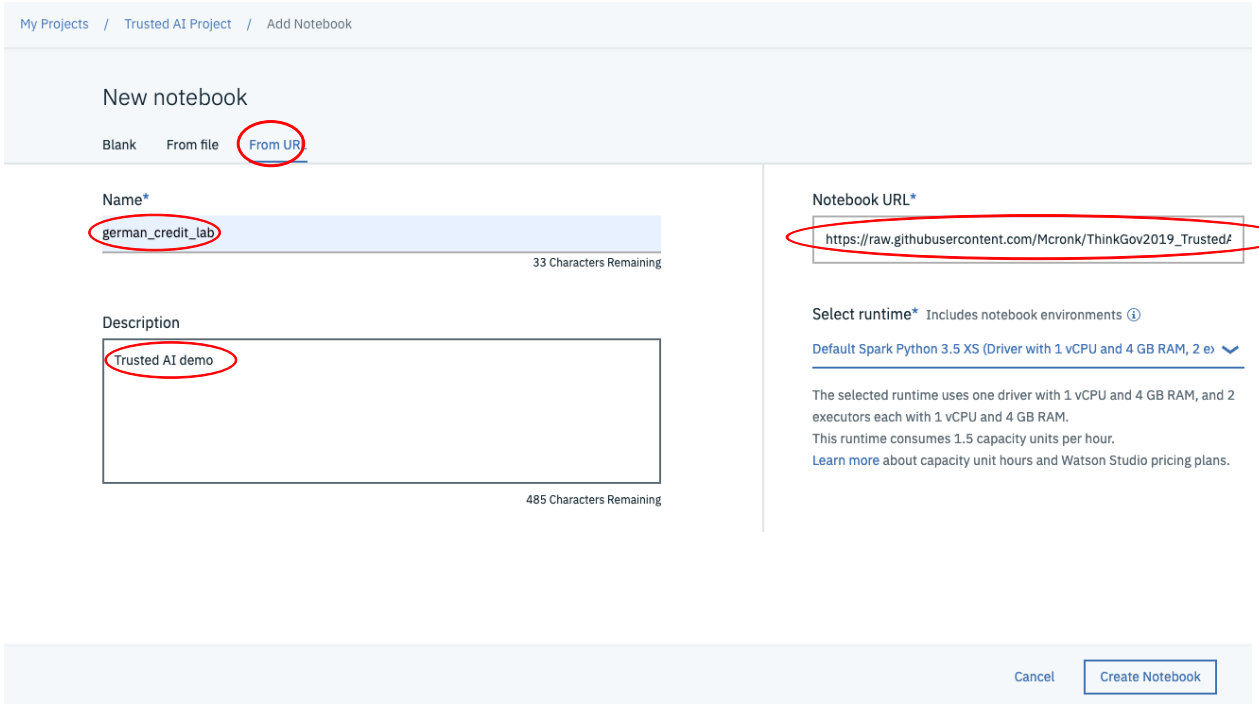
1. Click on Add to project.



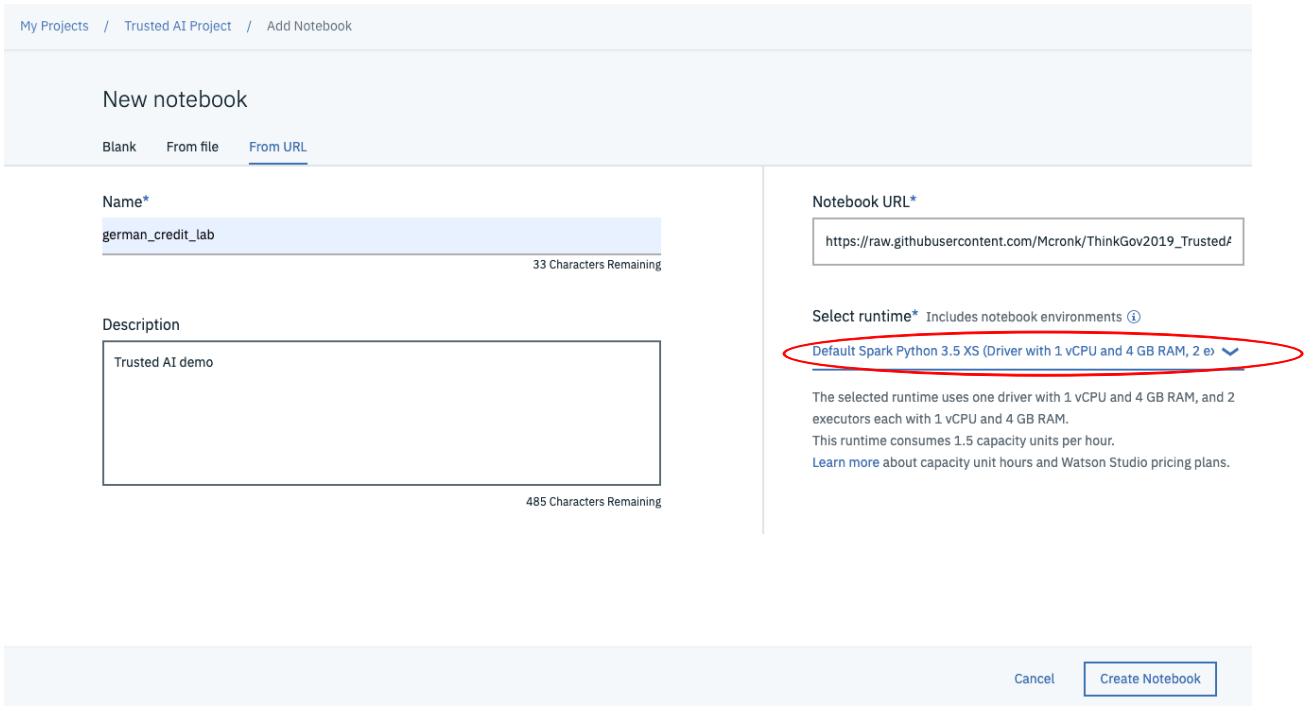
2. Click on Notebook.



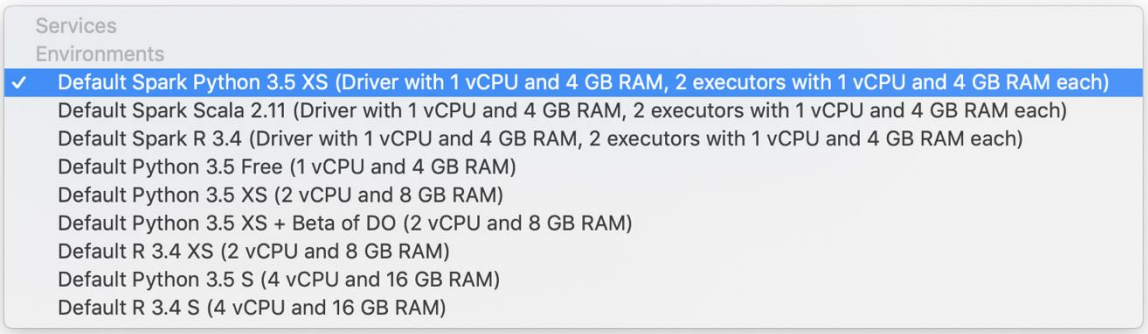
3. Click on the From URL tab.
4. Enter a Name and Description (optional) for your notebook.
5. Under Notebook URL, enter:
   https://raw.githubusercontent.com/Mcronk/ThinkGov2019_TrustedAI/master/german_credit_lab.ipynb



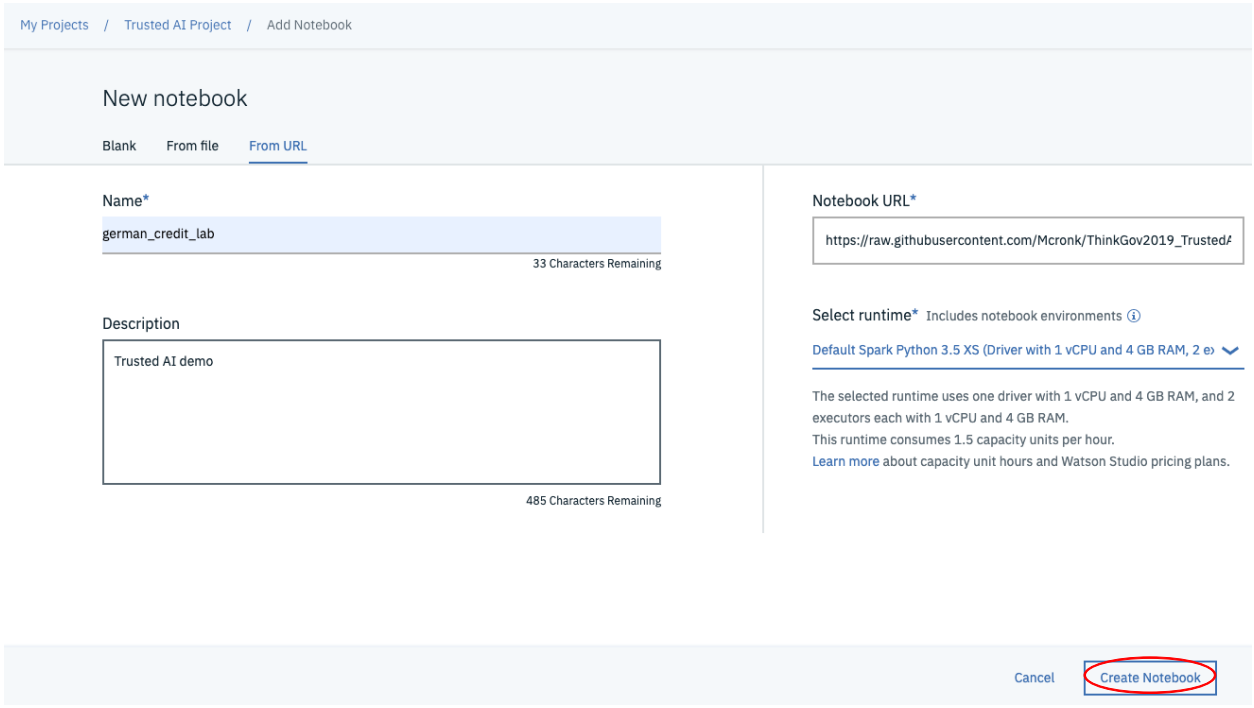
6. Under Select runtime, click on the dropdown menu.



7. Choose Default Spark Python 3.5 XS

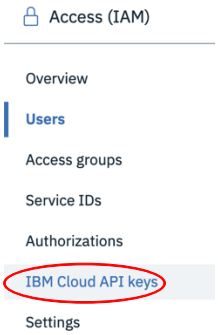


8. Click Create Notebook.

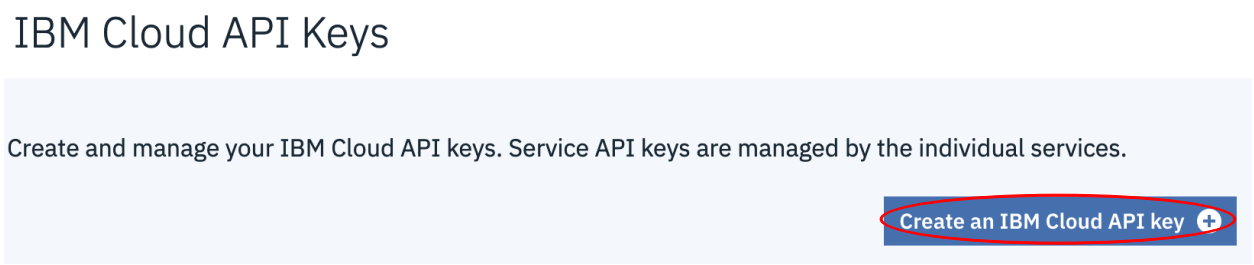


Our notebook will create and deploy a machine learning model then configure Watson OpenScale and connect it to our cloud machine learning service. Before running our notebook, we need to enter credentials for the IBM Cloud API and our machine learning service.

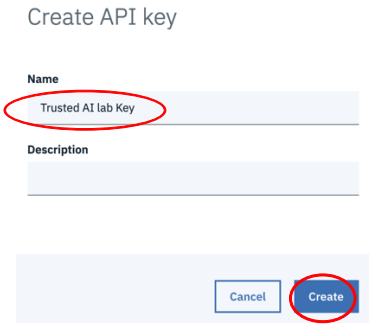
9. Open up a new browser tab and access https://cloud.ibm.com/iam#/users
10. On the left side of the screen, click on IBM Cloud API keys.



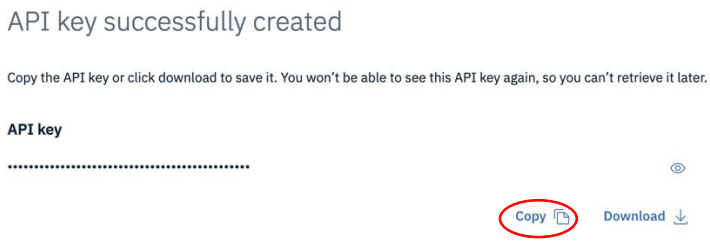
11. Along the center of your screen, click on Create a Cloud API Key.



12. Give the API key a name and click Create.

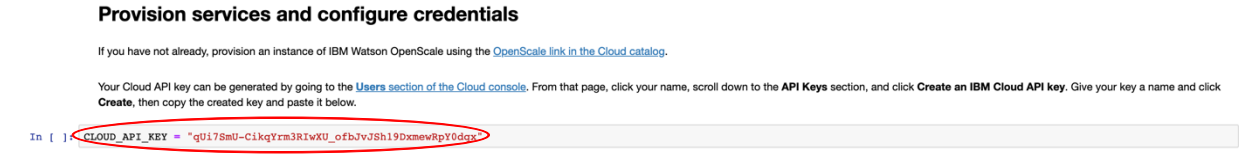


13. Click on Copy.

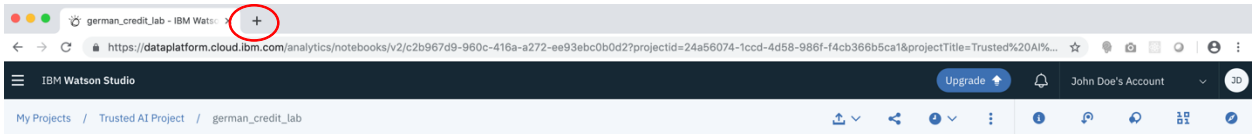


14. Back in your notebook under Provision service and configure credentials (next to CLOUD_API_KEY = , replace “PASTE HERE” with the key you just copied.

Your notebook should match the image below (with a different key between the quotes).

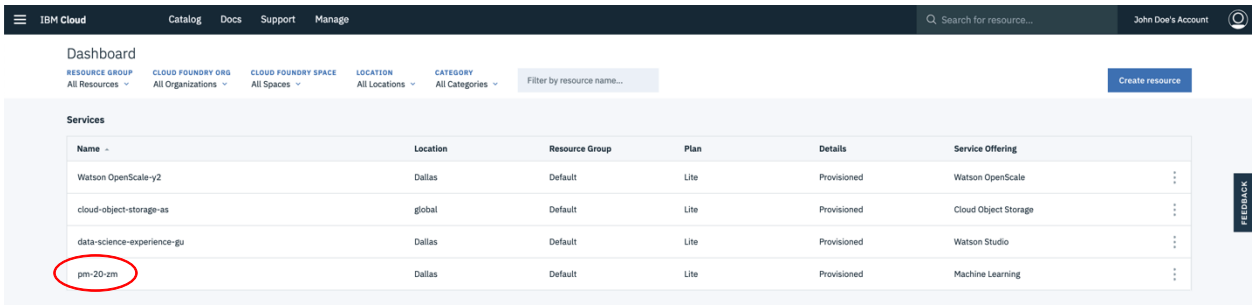


15. Open up a new browser tab and access bluemix.net



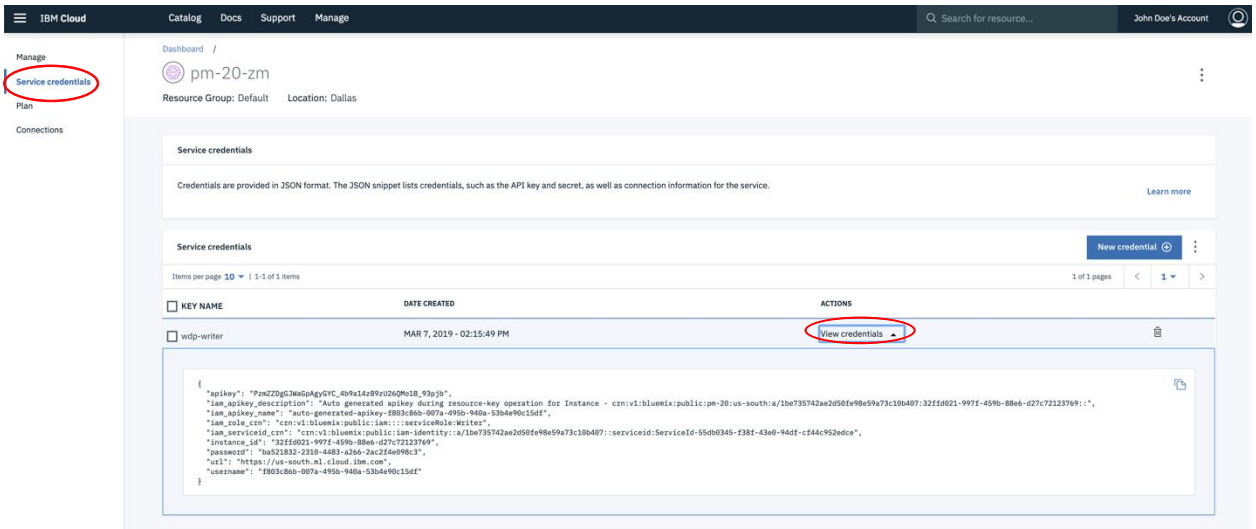
You will be brought to your dashboard page for Cloud resources.

16. Click on the Machine Learning service (the example in these instructions is called “pm-20-zm”

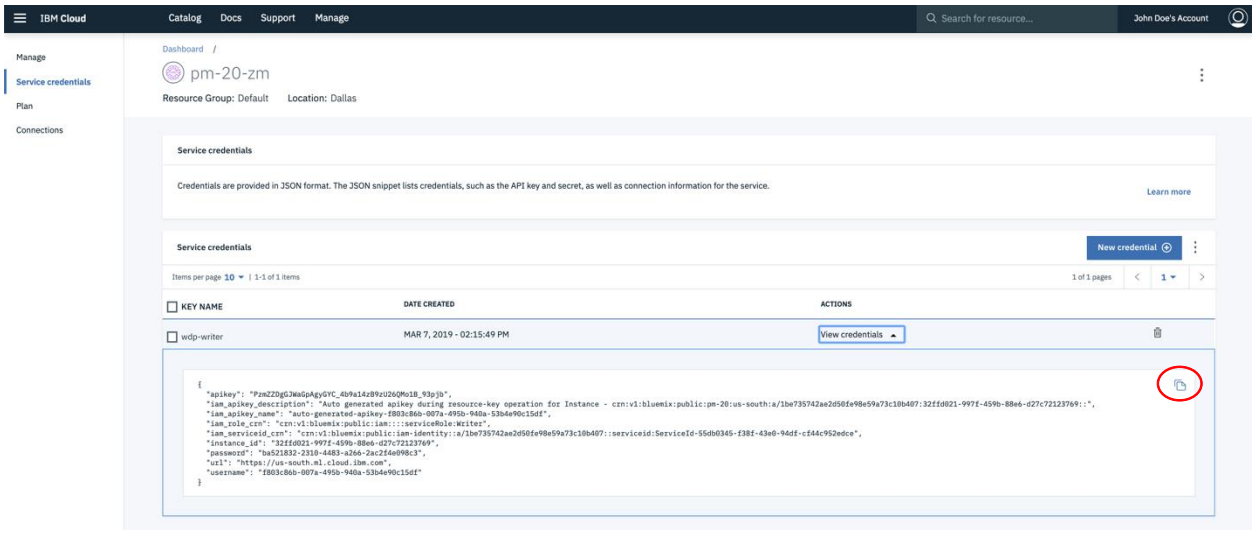


17. Click on Service Credentials along the left side.

18. Then click on View Credentials along the center.



19. Click on the pages icon to copy the credentials.



25. Back in your notebook, under Provision services and configure credentials, replace the WML_CREDENTIALS = brackets with what you have just copied. WML stands for Watson Machine Learning.

Provision services and configure credentials

If you have not already, provision an instance of IBM Watson OpenScale using the OpenScale link in the Clou

Your Cloud API key can be generated by going to the **Users** section of the Cloud console. From that page, cl
Create, then copy the created key and paste it below.

```
In [ ]: CLOUD_API_KEY = "qUi7SmU-CikqYrm3RIwXU_ofbJvJSh19DxmewRpY0dqx"
```

Next you will need credentials for Watson Machine Learning. If you already have a WML instance, you may us
Once your instance is created, click the **Service Credentials** link on the left side of the screen. Click the **New credentials** button. Copy and paste your WML credentials into the cell below.

```
In [ ]: WML_CREDENTIALS = {
            "apikey": "key",
            "iam_apikey_description": "description",
            "iam_apikey_name": "auto-generated-apikey",
            "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",
            "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::",
            "instance_id": "instance_id",
            "password": "password",
            "url": "https://us-south.ml.cloud.ibm.com",
            "username": "username"
        }
```

Your notebook should match the image below (with different credentials).

Provision services and configure credentials

If you have not already, provision an instance of IBM Watson OpenScale using the OpenScale link in the Cloud catalog

Your Cloud API key can be generated by going to the **Users** section of the Cloud console. From that page, click your n
Create, then copy the created key and paste it below.

```
In [ ]: CLOUD_API_KEY = "qUi7SmU-CikqYrm3RIwXU_ofbJvJSh19DxmewRpY0dqx"
```

Next you will need credentials for Watson Machine Learning. If you already have a WML instance, you may use creden
Once your instance is created, click the **Service Credentials** link on the left side of the screen. Click the **New credenti
credentials button. Copy and paste your WML credentials into the cell below.

```
In [ ]: WML_CREDENTIALS = {
  "apikey": "PzmZZDgGJWaGpAgyGYC_4b9a14zB9zU26QMo1B_93pjb",
  "iam_apikey_description": "Auto generated apikey during resource-key operation for Insta
  "iam_apikey_name": "auto-generated-apikey-f803c86b-007a-495b-940a-53b4e90c15df",
  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",
  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a/1be735742ae2d50fe98e59a73c10
  "instance_id": "32ffd021-997f-459b-88e6-d27c72123769",
  "password": "ba521832-2310-4483-a266-2ac2f4e098c3",
  "url": "https://us-south.ml.cloud.ibm.com",
  "username": "f803c86b-007a-495b-940a-53b4e90c15df"
}
```

Our notebook now has all the required credentials.

20. Select each cell in the notebook and hit SHIFT+ENTER to run.

Run your notebook cell by cell. Read through the notebook and observe the output. With the exception of the first cell installing packages, there should be no error output.

Identify transactions for Explainability

Transaction IDs identified by the cells below can be copied and pasted into the Explainability tab of the OpenScale dashboard.

```
In [64]: payload_data = subscription.payload_logging.get_table_content(limit=60)
         payload_data.filter(items=['scoring_id', 'predictedLabel', 'probability'])
```

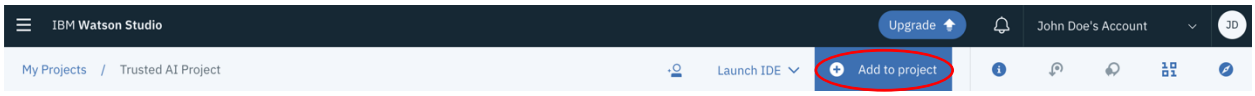
Out[64]:

	scoring_id	predictedLabel	probability
0	e42459305c0c44dc185c6bc144d74806-423	No Risk	[0.8128085875806539, 0.18719141241934611]
1	e42459305c0c44dc185c6bc144d74806-567	No Risk	[0.9576935217341755, 0.0423064782658244]
2	e42459305c0c44dc185c6bc144d74806-951	No Risk	[0.9509147179413443, 0.04908528205865579]
3	e42459305c0c44dc185c6bc144d74806-663	No Risk	[0.9565610290260815, 0.04343897097391844]
4	e42459305c0c44dc185c6bc144d74806-424	No Risk	[0.81396200025809, 0.18603799974191007]
5	e42459305c0c44dc185c6bc144d74806-568	No Risk	[0.5729289479961321, 0.4270710520038678]
6	e42459305c0c44dc185c6bc144d74806-1	No Risk	[0.7093257286263136, 0.29067427137368634]
7	e42459305c0c44dc185c6bc144d74806-807	No Risk	[0.5765253746425008, 0.42347462535749925]
8	e42459305c0c44dc185c6bc144d74806-425	No Risk	[0.7136291870129947, 0.28637081298700523]
9	e42459305c0c44dc185c6bc144d74806-569	No Risk	[0.5485131522852063, 0.4514868477147938]
10	e42459305c0c44dc185c6bc144d74806-952	Risk	[0.09948668541411484, 0.9005133145858851]

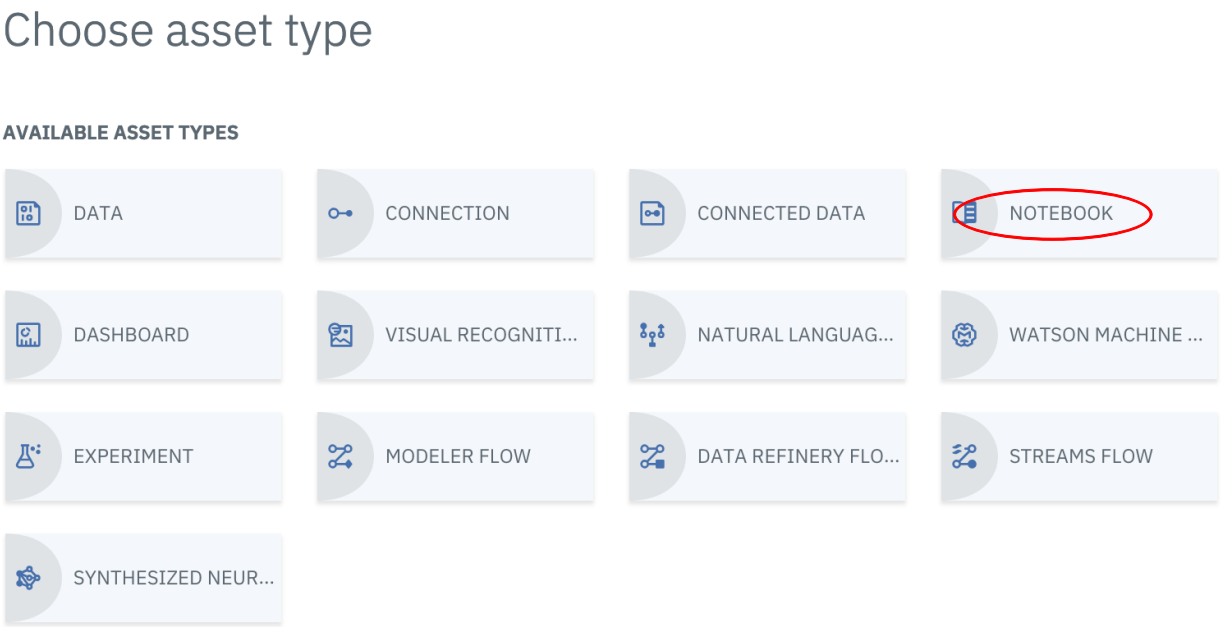
Once the last cell has run successfully, continue with the steps below.

We will now run a notebook that will automatically feed data to our deployed ML model to showcase transaction-level bias mitigation.

21. Return to your project dashboard.
22. Towards the top of your screen, click on Add to project.

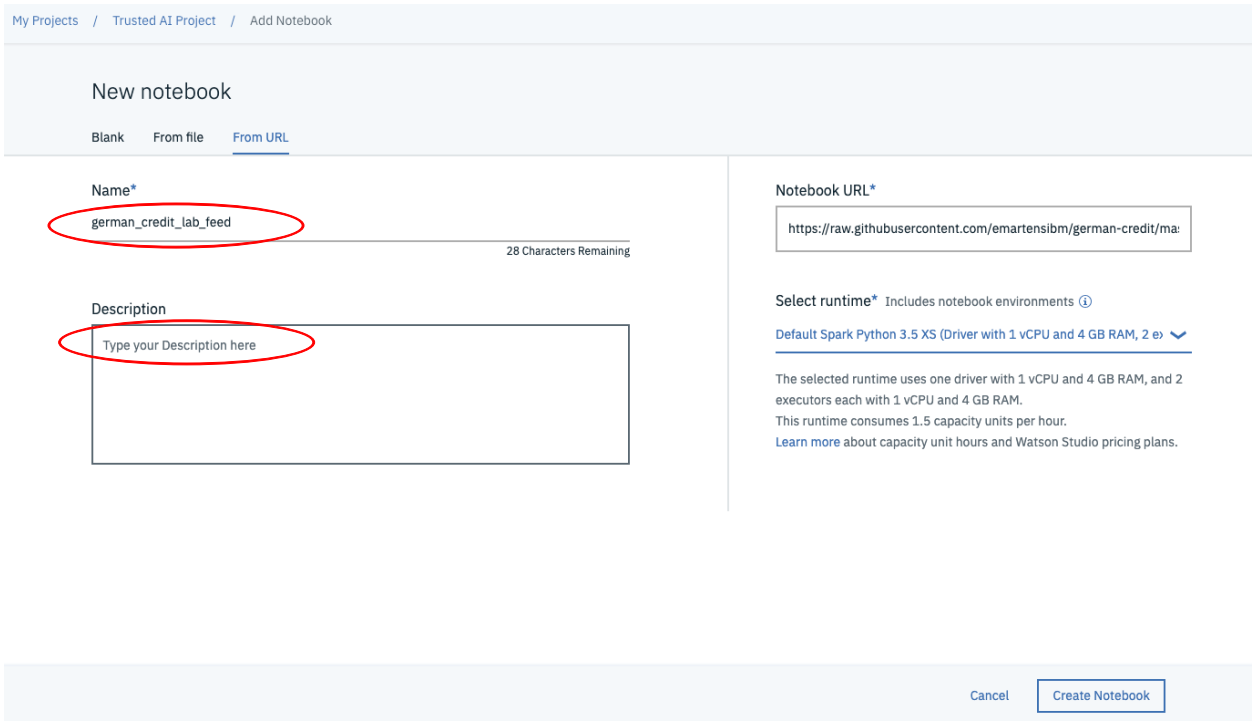


23. Click on Notebook.

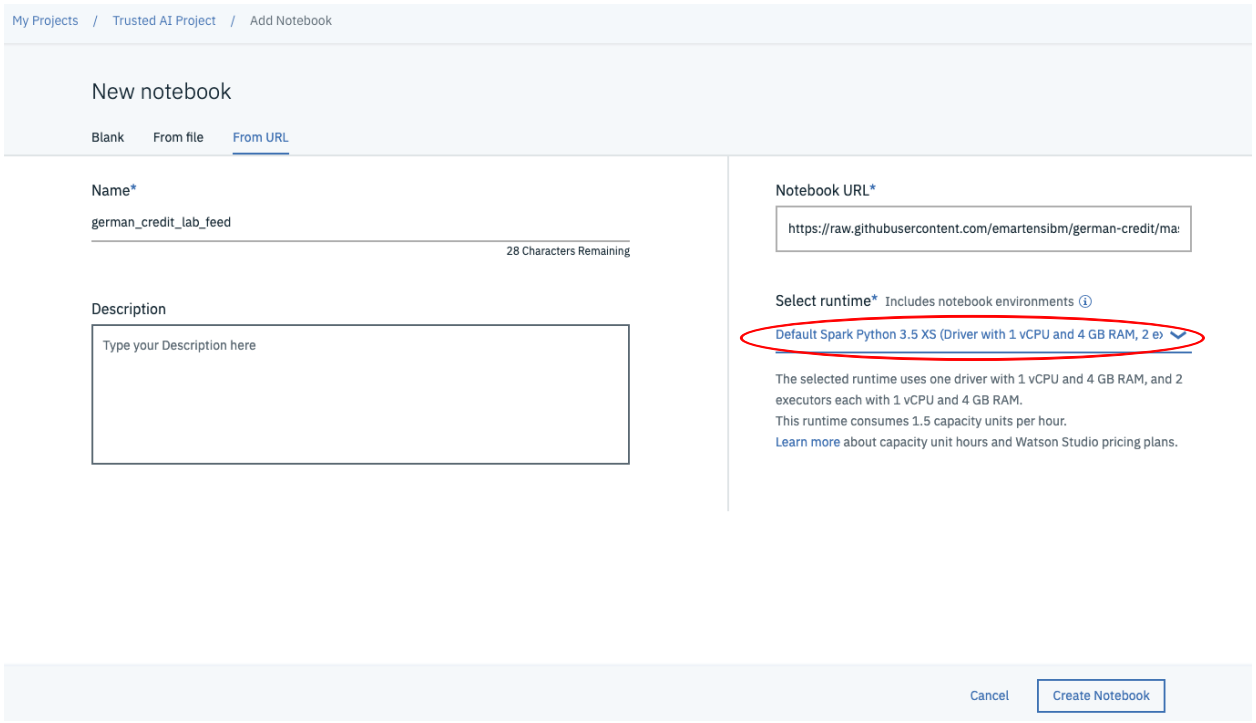


24. Enter a Name and Description for your notebook.
25. Under Notebook URL, enter:

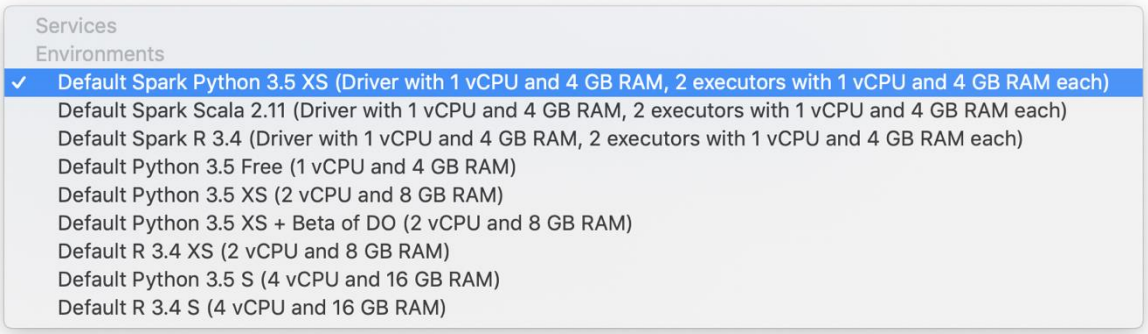
https://raw.githubusercontent.com/emartensibm/german-credit/master/german_credit_scoring_feed.ipynb



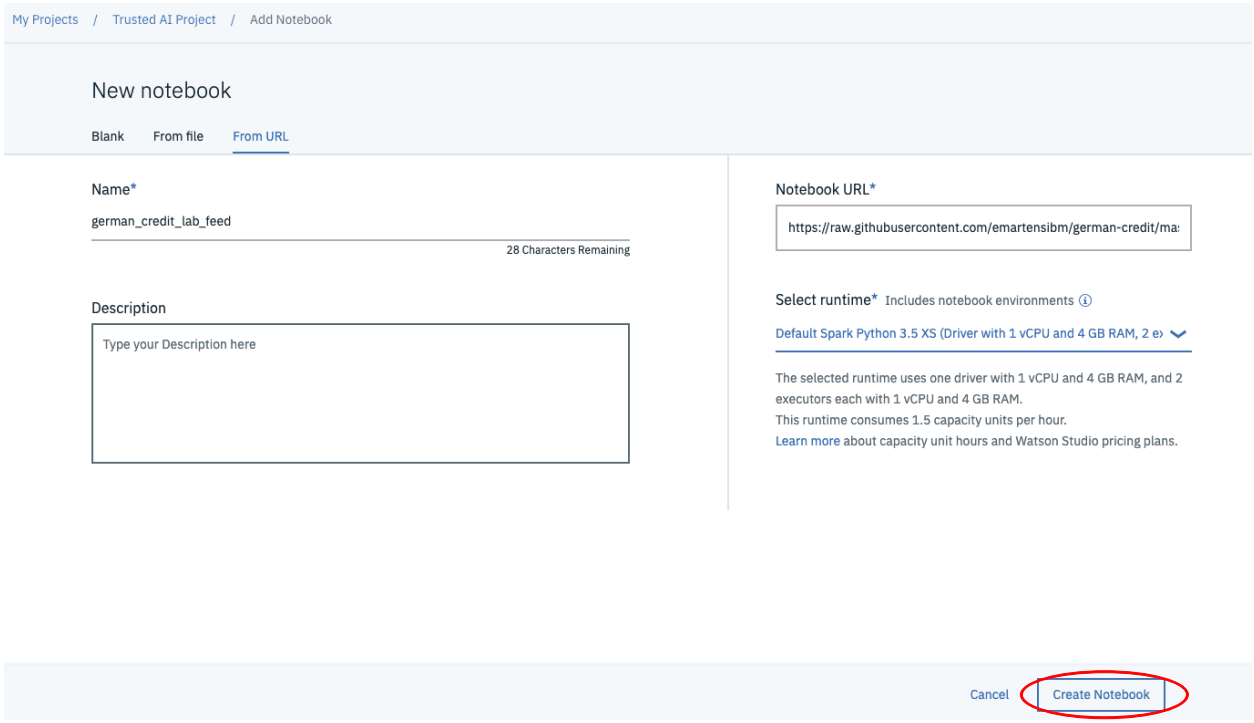
26. Under Select runtime, click on the dropdown menu.



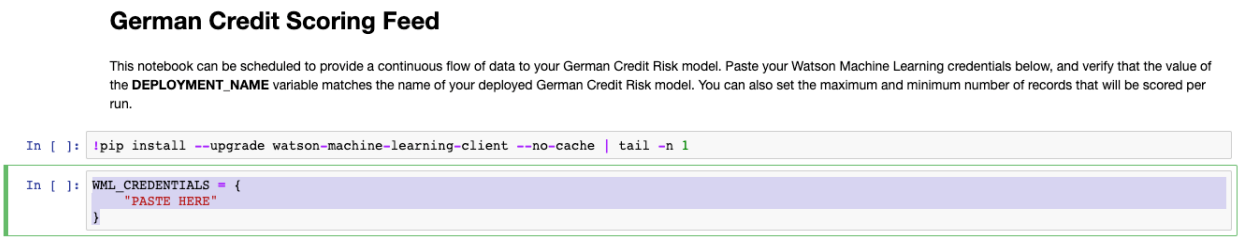
27. Choose Default Spark Python 3.5 XS



28. Click Create Notebook.



29. As in the previous notebook, enter your Watson Machine Learning (WML) credentials in the second code box.



German Credit Scoring Feed

This notebook can be scheduled to provide a continuous flow of data to your German Credit Risk model. Paste your Watson Machine Learning credentials below, and verify that the value of the **DEPLOYMENT_NAME** variable matches the name of your deployed German Credit Risk model. You can also set the maximum and minimum number of records that will be scored per run.

In [ ]:
```
!pip install --upgrade watson-machine-learning-client --no-cache | tail -n 1
```

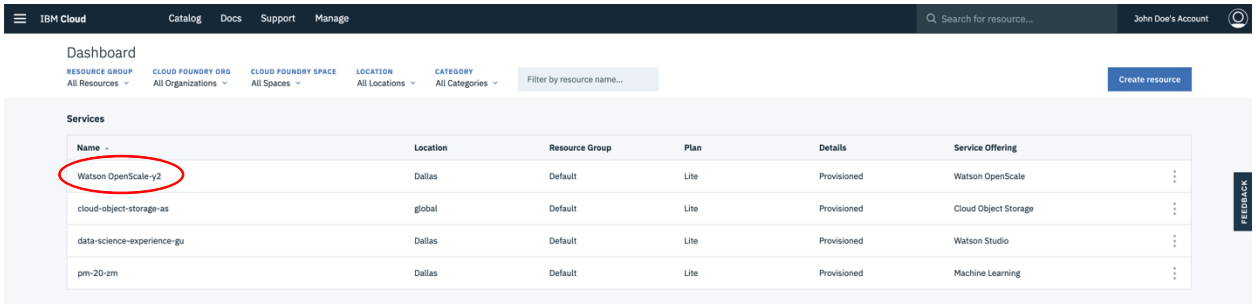
In [ ]:
```
WML_CREDENTIALS = {
  "apikey": "gfVRg_Oyql9DpzmeH9101z1kN02qi3FbyVg5iwZbDTf7",
  "iam_apikey_description": "Auto generated apikey during resource-key operation for Instance - crn:v1:bluemix:public:pm-20:us-south:a/1be735742a
  "iam_apikey_name": "auto-generated-apikey-9d67e9ea-2f9c-408f-b738-513e65618c8c",
  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",
  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a/1be735742ae2d50fe98e59a73c10b407::serviceid:ServiceId-9311cb91-d07b-457c-ad1e-8fe81
  "instance_id": "a04620c8-926c-480c-9e8f-563e74c4481a",
  "password": "16befb2e-c614-4d01-be64-efdc9554601b",
  "url": "https://us-south.ml.cloud.ibm.com",
  "username": "9d67e9ea-2f9c-408f-b738-513e65618c8c"
}
```

30. Run the notebook cell by cell.

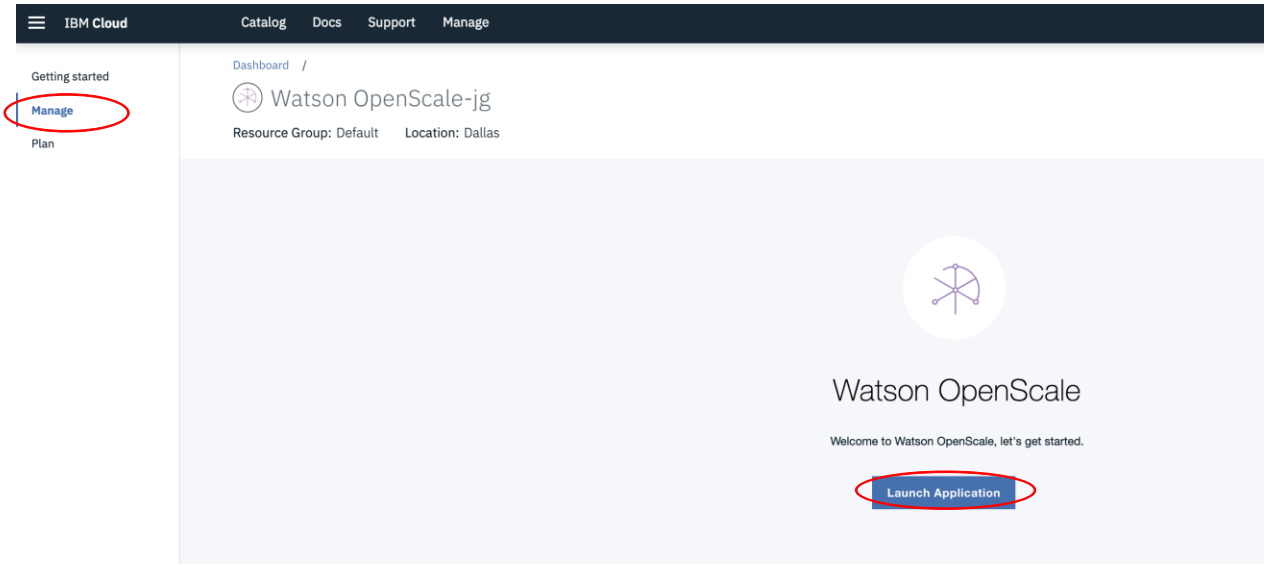
Once the notebook has completed, move on to the next step.

Step 4: Explore Watson OpenScale.
For the purpose of this lab we have used notebooks to showcase the API and speed along the configuration process. For documentation and tutorials on manual configuration, please visit https://cloud.ibm.com/docs/services/ai-openscale?topic=ai-openscale-crt-ov#crt-ov

1. Return to https://console.bluemix.net/dashboard/apps
2. Click on the name of your Watson OpenScale service.

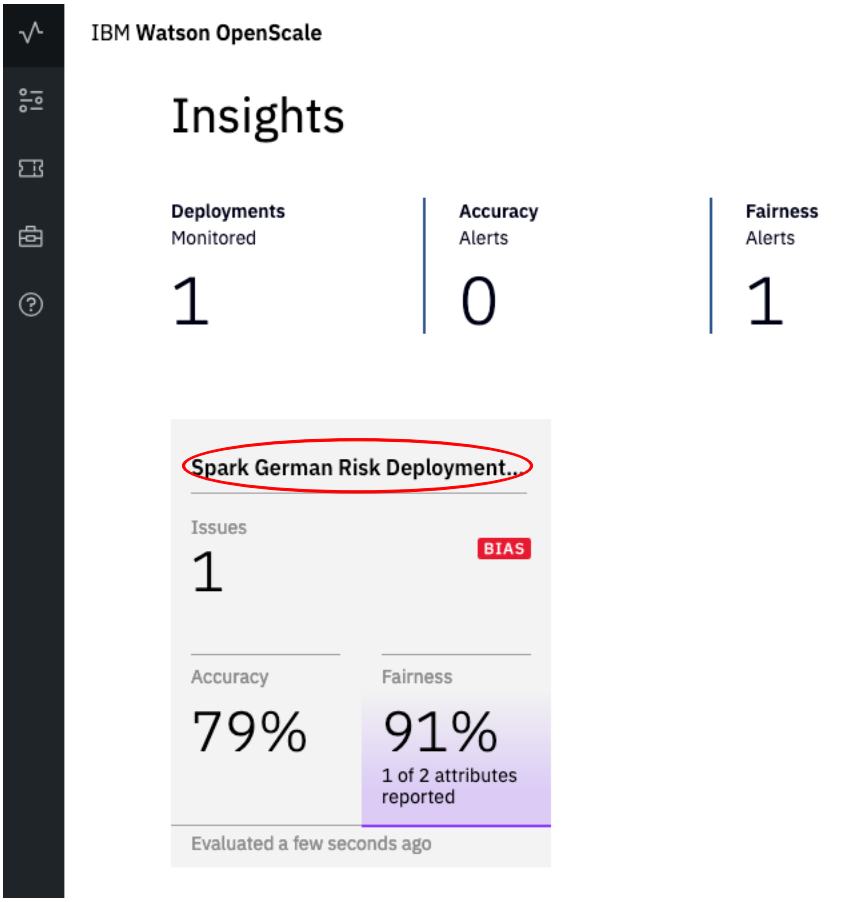


3. Click on Manage along the bar on the left side.
4. Click on Launch Application.



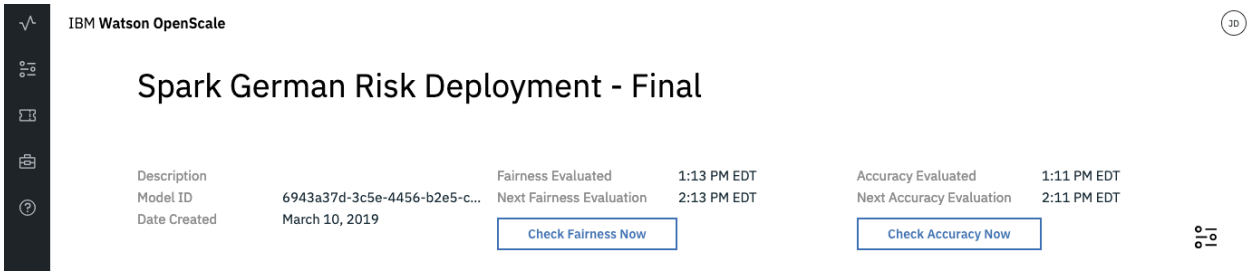
You will be brought to your Watson OpenScale dashboard. This is meant to show summary information on all models connected to this instance of Watson OpenScale. At a glance we are able to check all our models for accuracy, fairness, and if there is any bias detected.

5. Click on your model card.



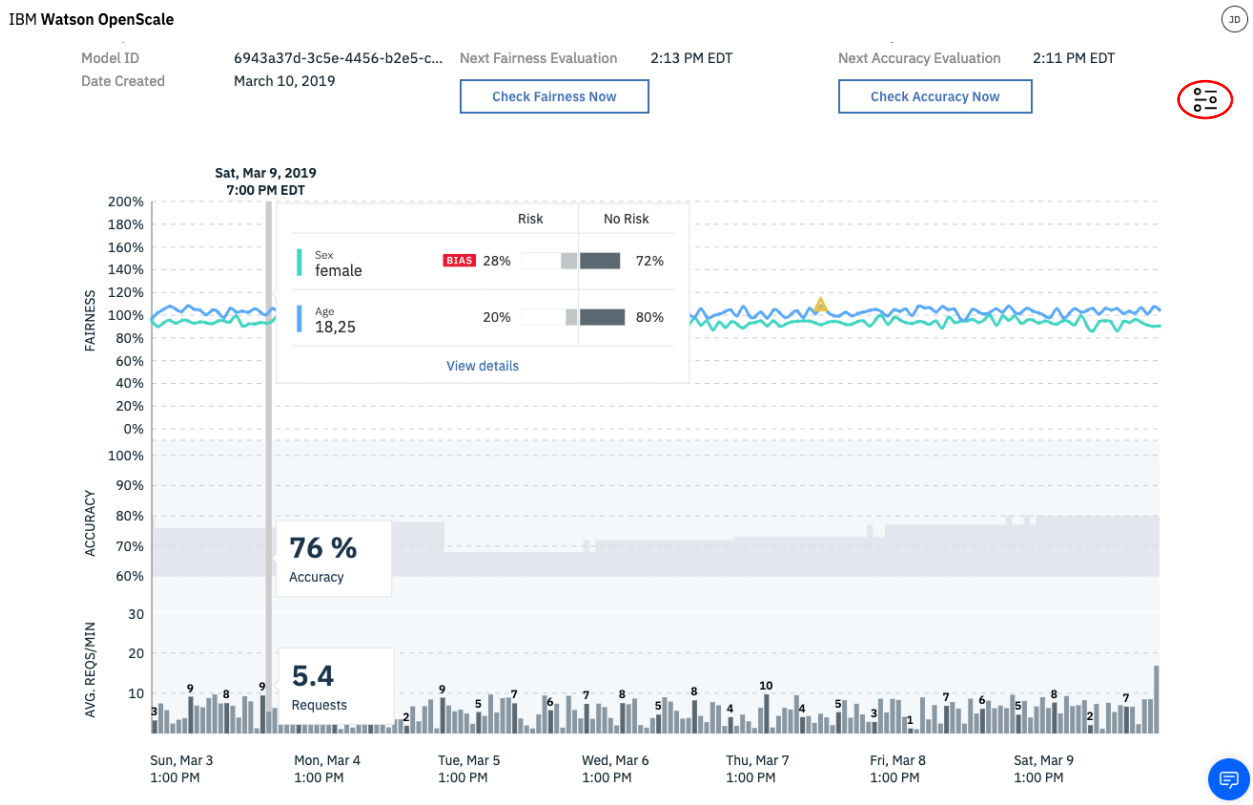
You will be brought to a model page.

The top of the screen shows model information and has buttons to immediately recalculate fairness and accuracy.



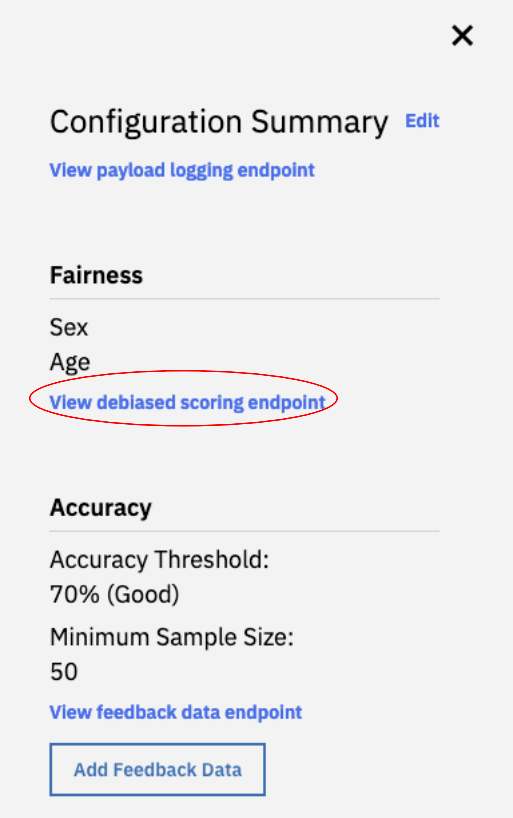
The lower half of the screen shows a time based graph tracking average requests per minute, accuracy, and fairness. Hovering over any part of the graph will allow you to drill down and see details for that set of transactions.

6. Click on the three dots to the far right of the screen.



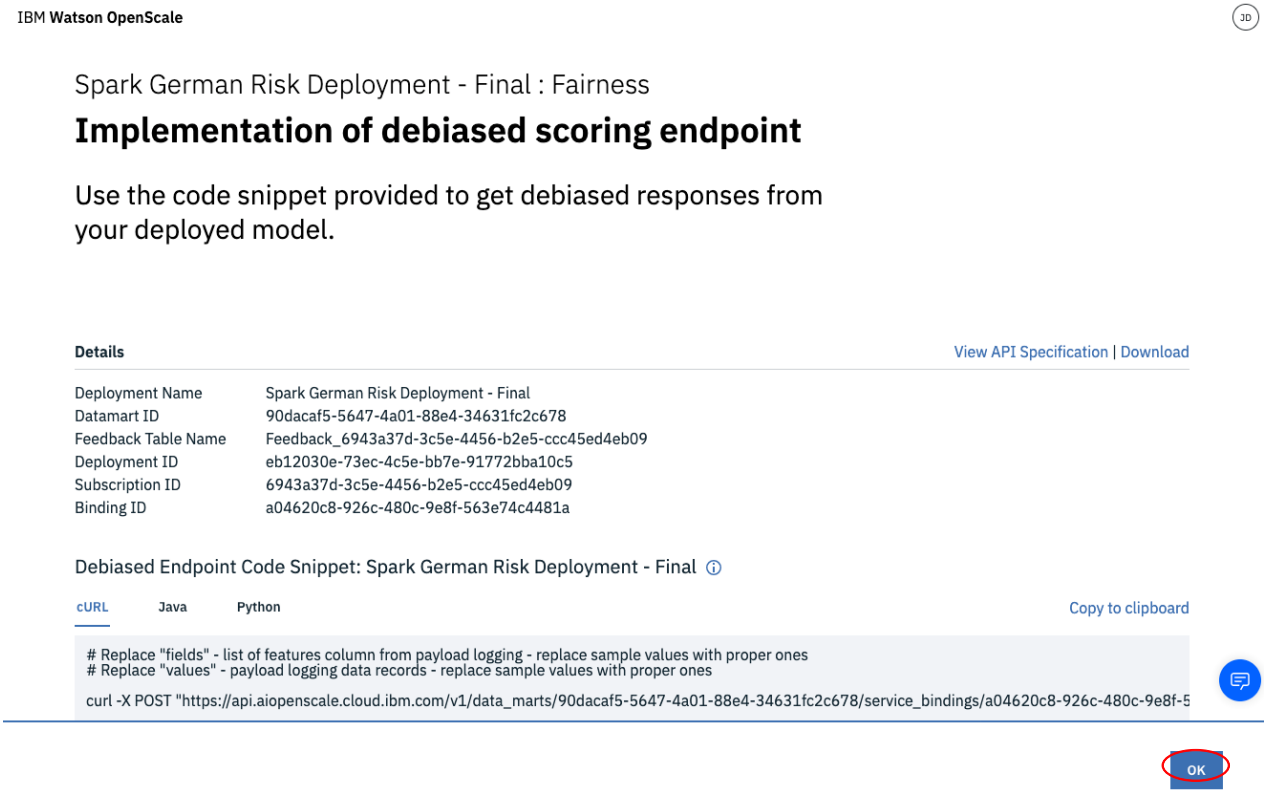
A sidebar will open up that has links to manually configure Watson OpenScale fairness/bias thresholds and tracked features/inputs as well as links to the debiased scoring endpoint for your model.

7. Click on View debiased scoring endpoint.

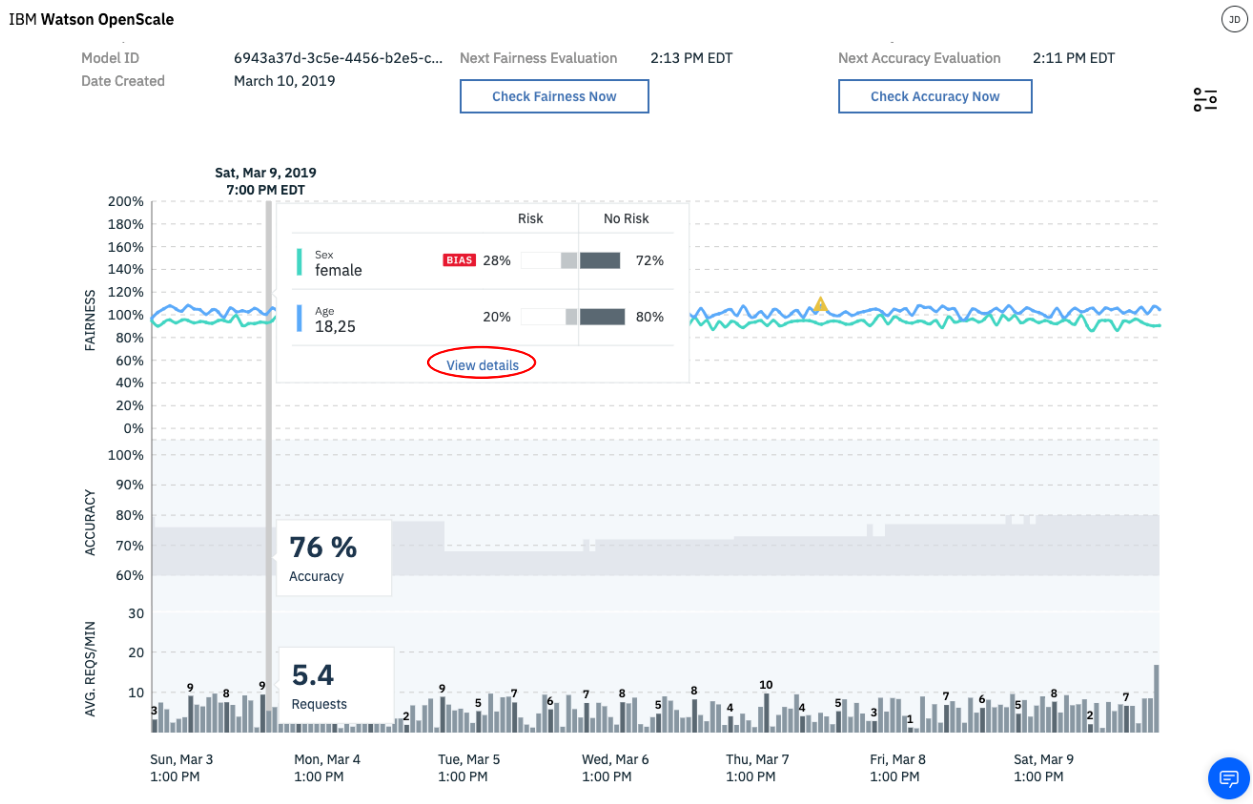


This provides API specifications and calls for an endpoint that automatically mitigates any bias detected in your machine learning model. This allows external systems to use an unbiased version of a deployed machine learning model on a transaction by transaction basis.

8. Click on OK to return to your model graph.



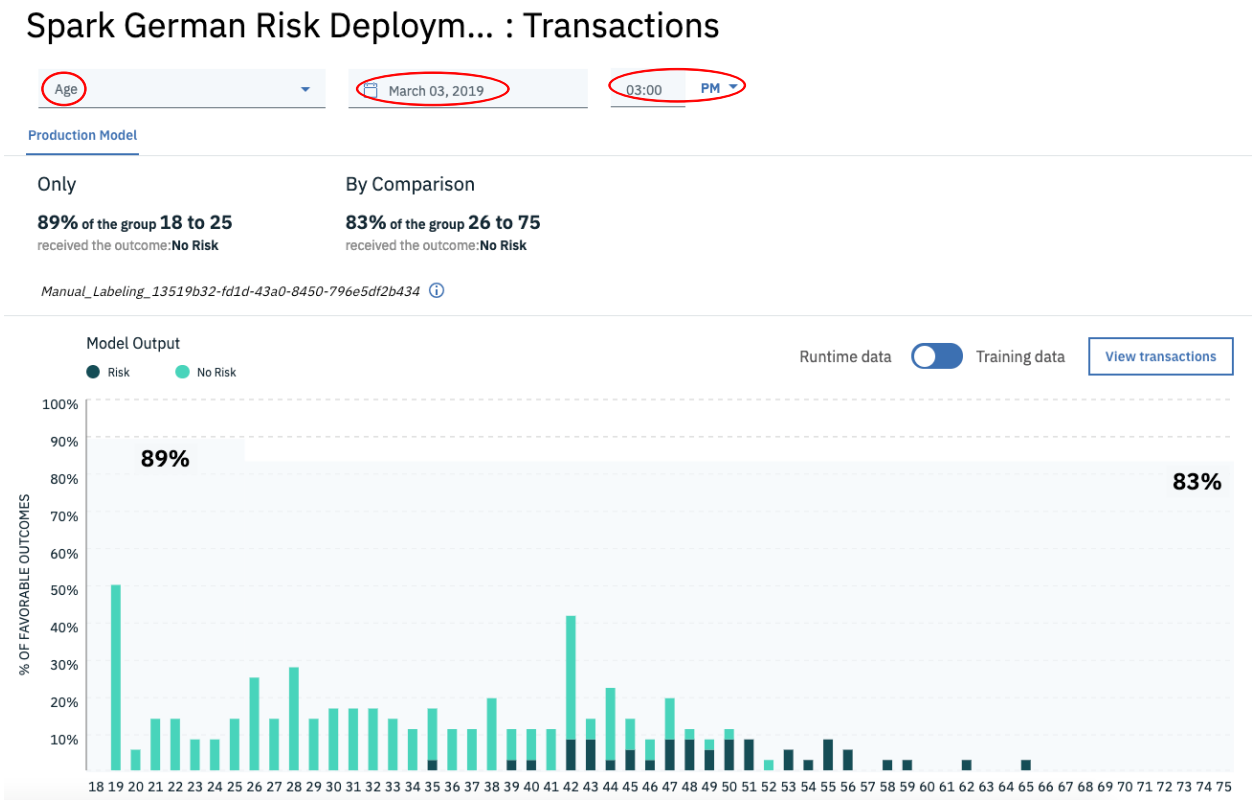
9. Hover over a more recent transaction and click View details.



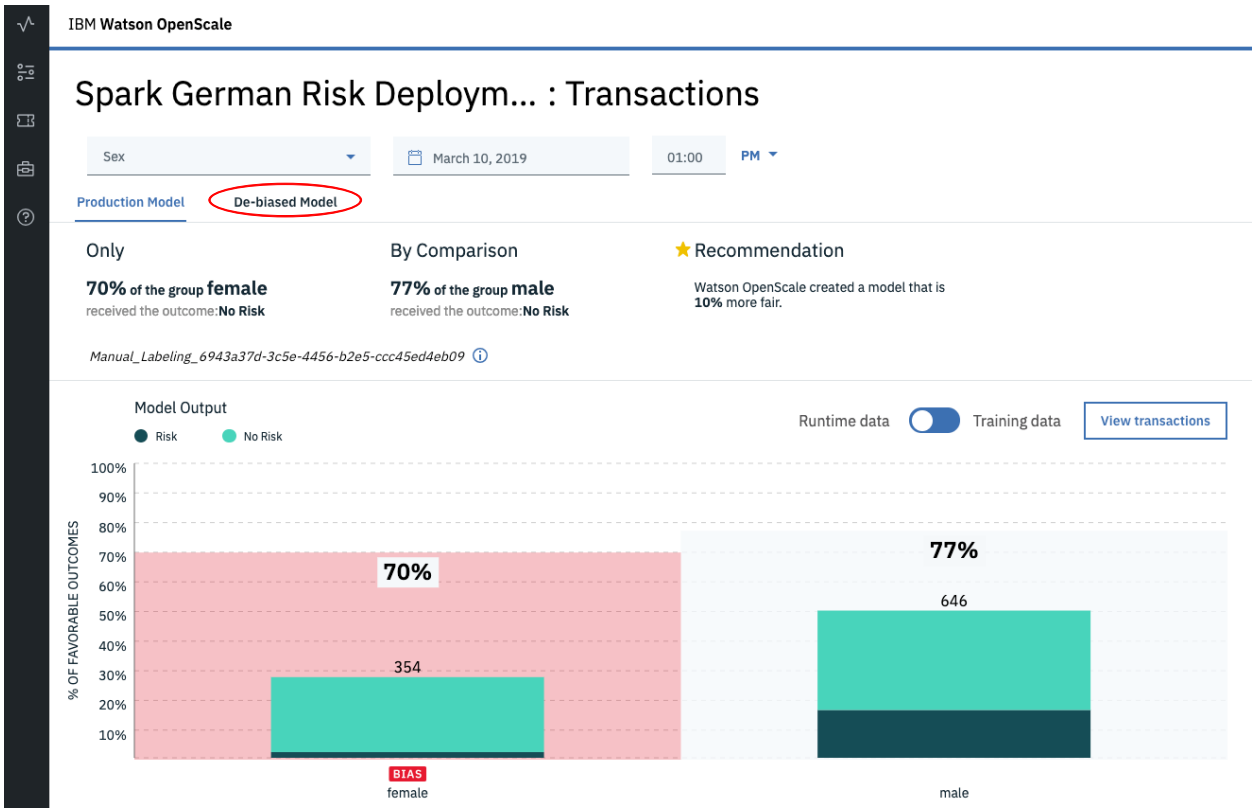
This page allows us to focus on fairness for a specific input feature (in this case we have set OpenScale to track the fairness of Age and Sex). Favorable outcomes were also set by our notebook with No Risk as the favorable outcome. A threshold can be configured to flag bias if fairness falls below that threshold.

Note that it may take 15 minutes time for the de-biased model to appear the first time.

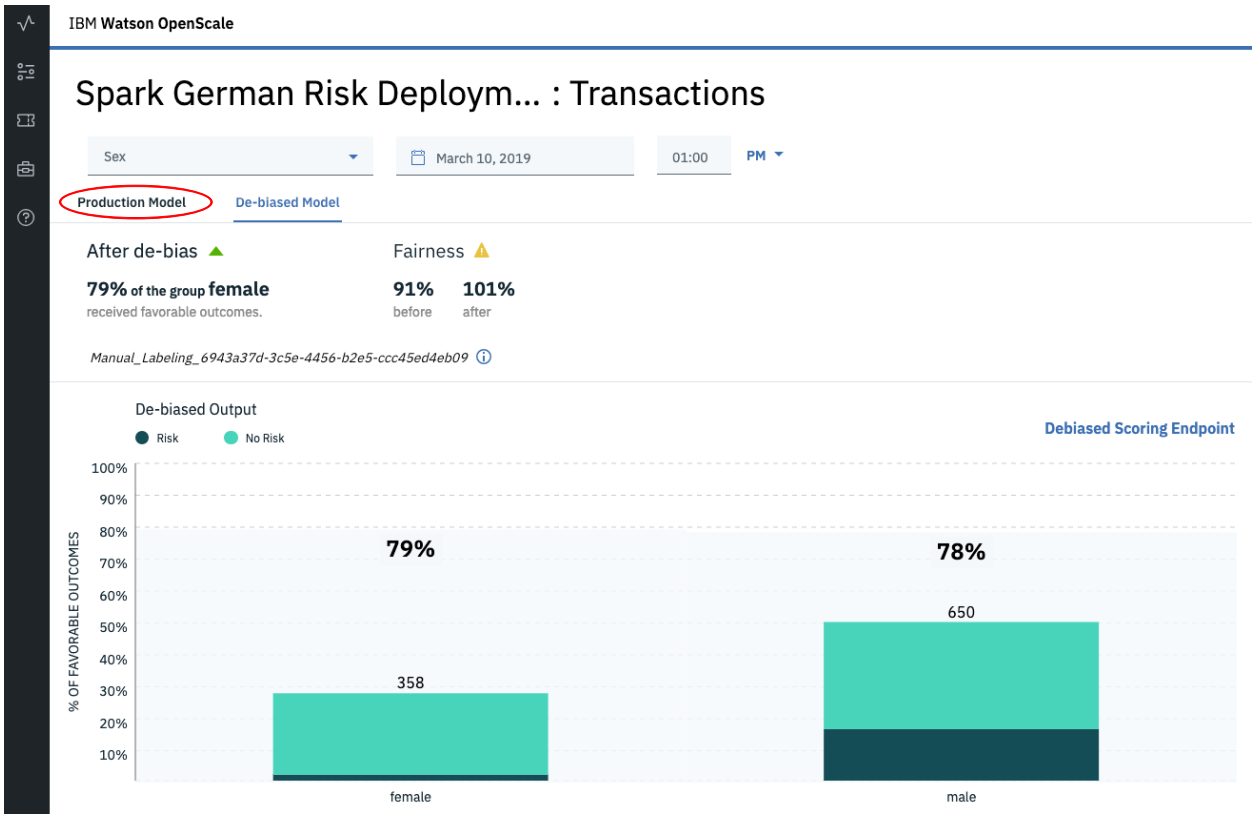
10. After waiting for 15 minutes, change Age to Sex.
11. Switch the date and time as close as possible to your current day and time.



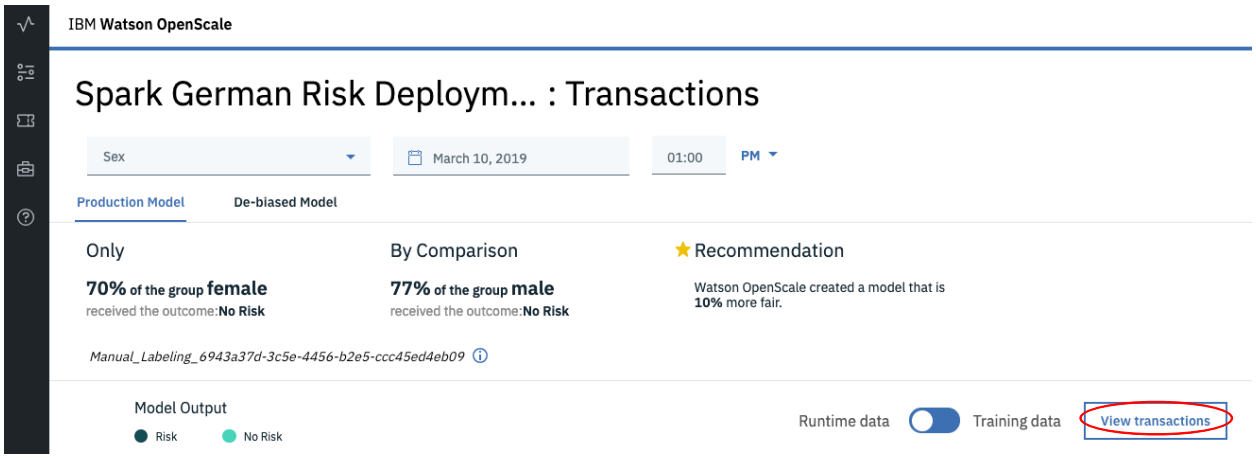
12. Switch to the De-biased model tab to observe the model OpenScale created to mitigate bias.



13. Once you have observed the change, click on the Production Model tab.

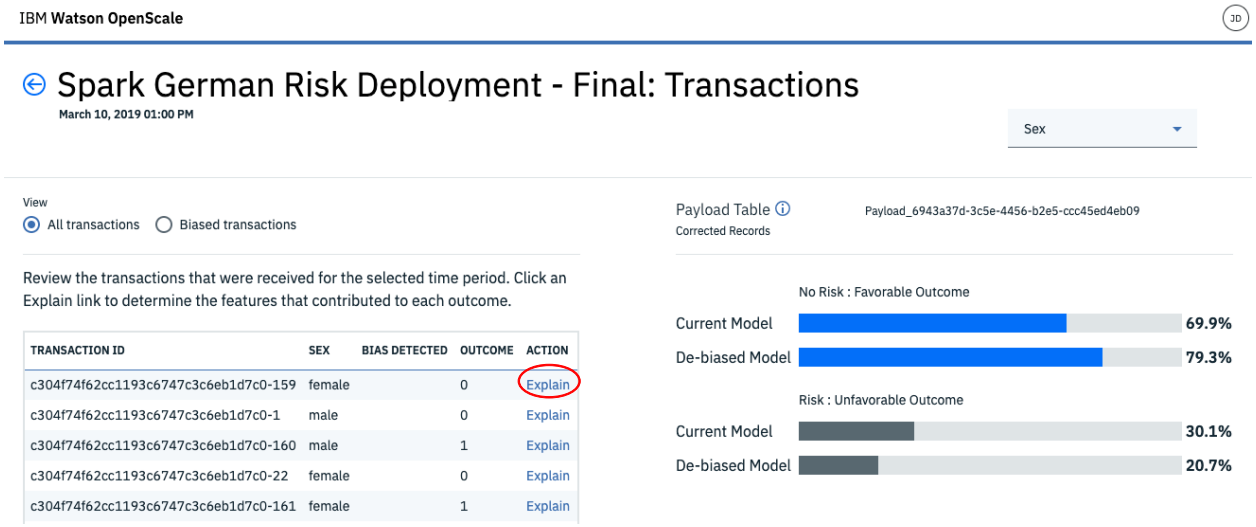


14. Click on View transactions.

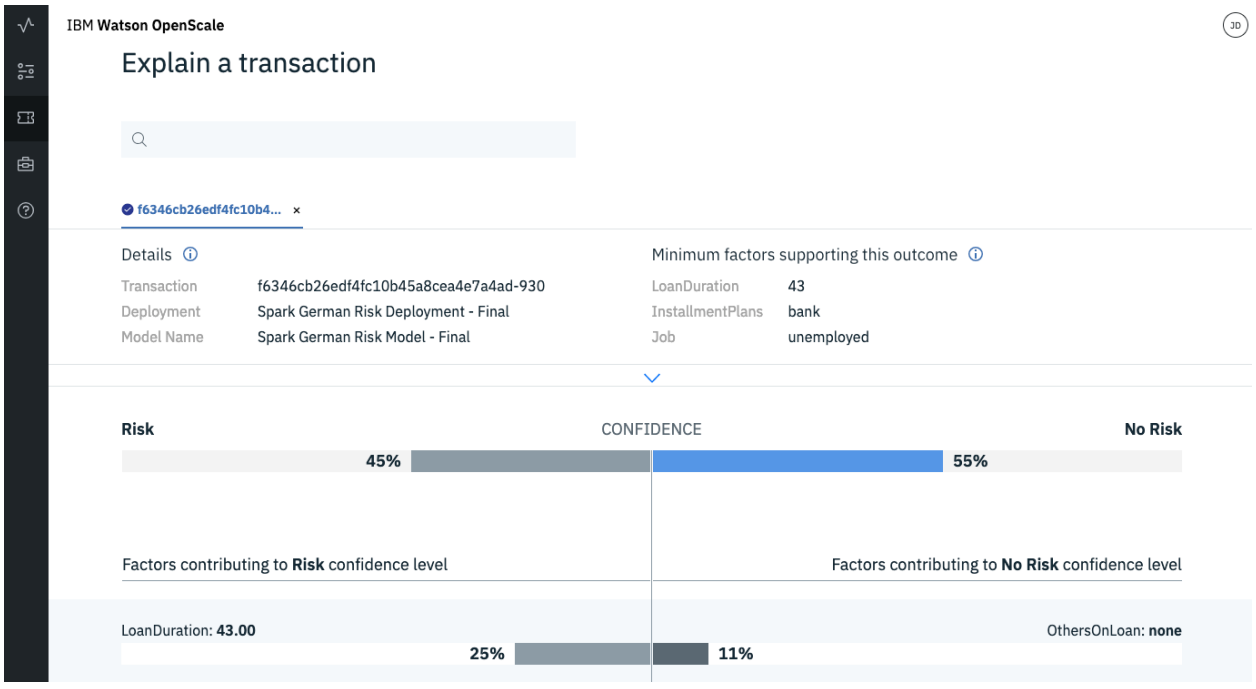


This page allows us to observe transaction-by-transaction lists on the left side of the screen and aggregate Payloads on the right with comparisons between the current model and de-biased models.

15. Click on Explain next to a transaction.



This shows how our model made a decision based on how different input features were weighted and lists the minimum changes necessary to switch the model decision.



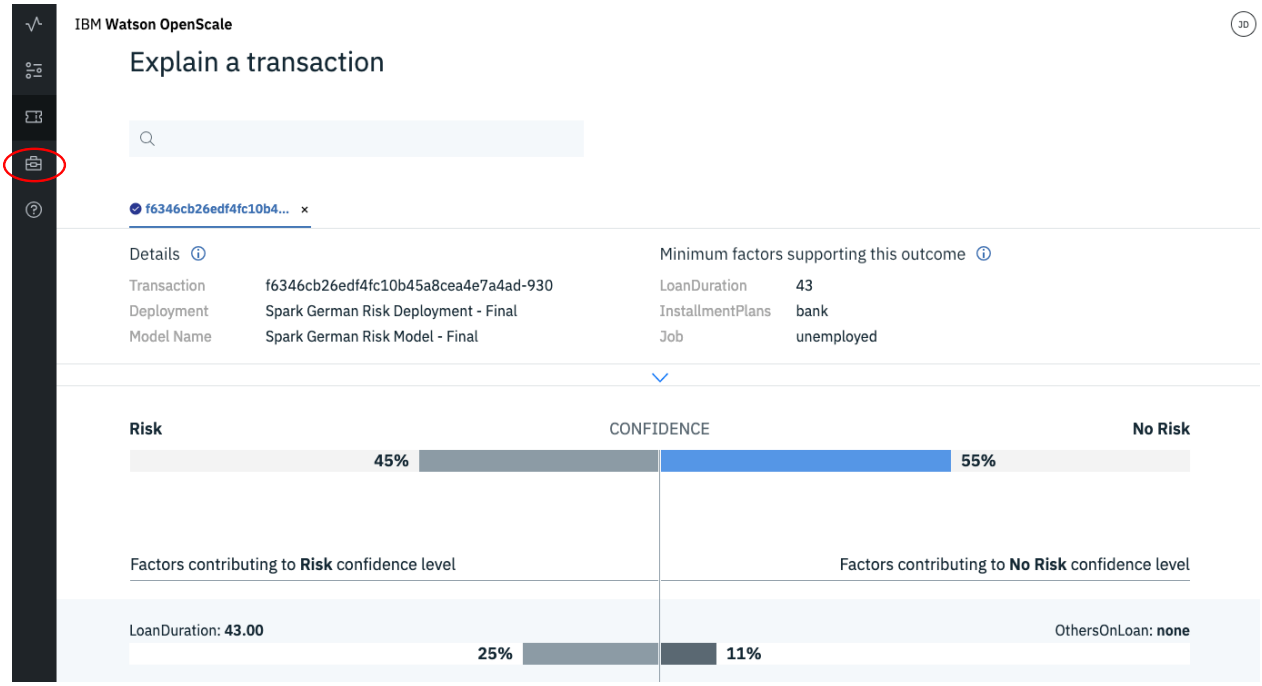
You have completed base lab.

For training on topics including machine learning, data science, and blockchain, please visit: https://www.ibm.com/industries/federal/analytics

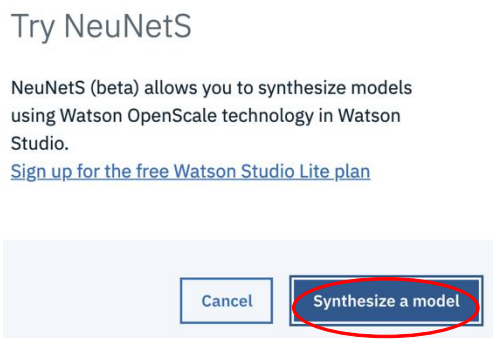
For an additional optional lab, continue on to step 4 on the next page.

Step 5: Synthesize a Neural Network. (Optional)
Watson OpenScale and Watson Studio are able to automatically synthesize Neural Networks with a high a degree of performance.

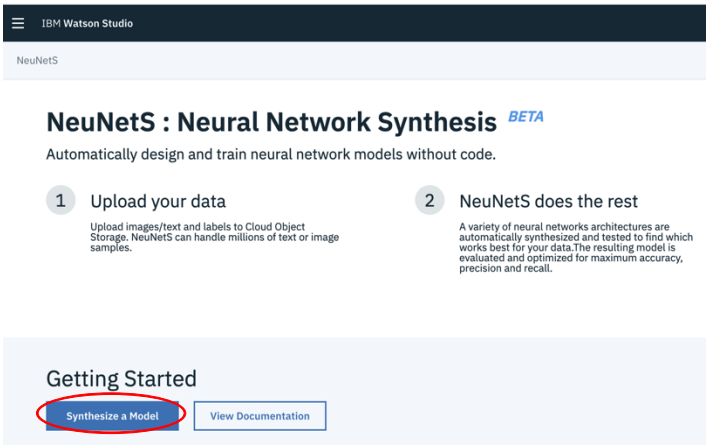
1. Click on the briefcase icon.



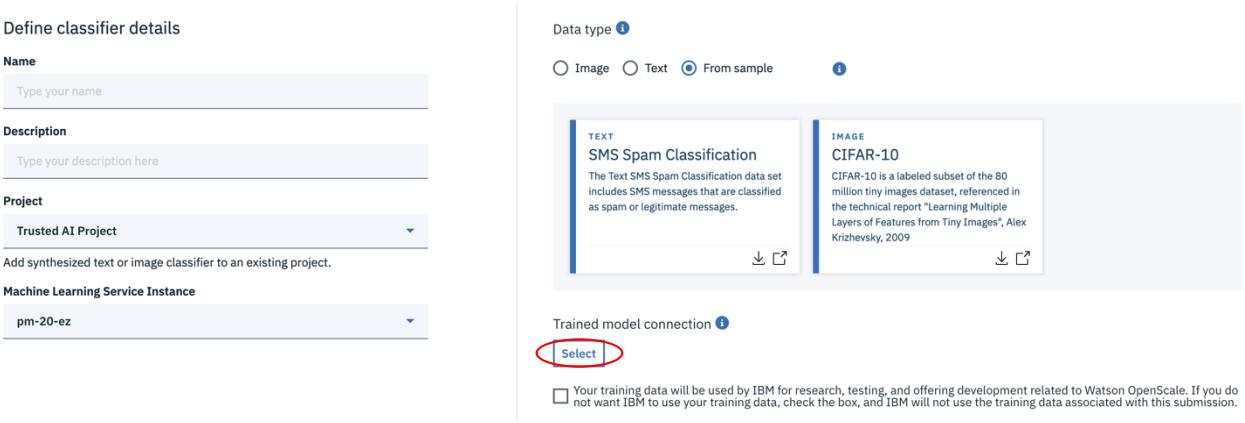
2. Click on Synthesize a model.



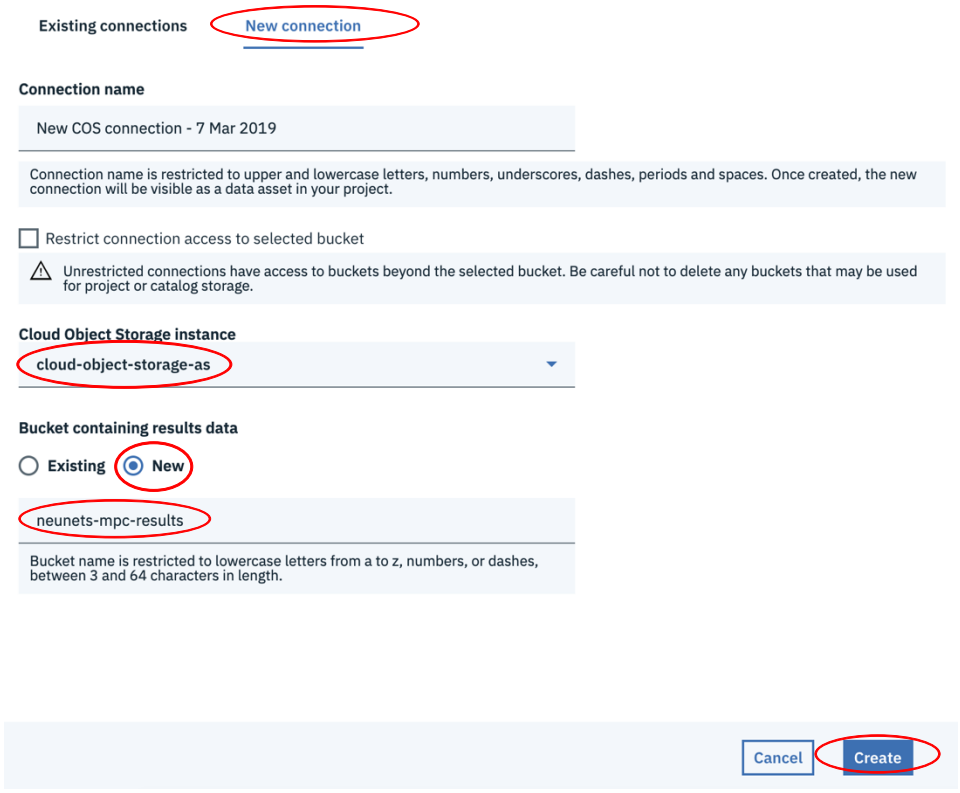
3. Click on Synthesize a model.



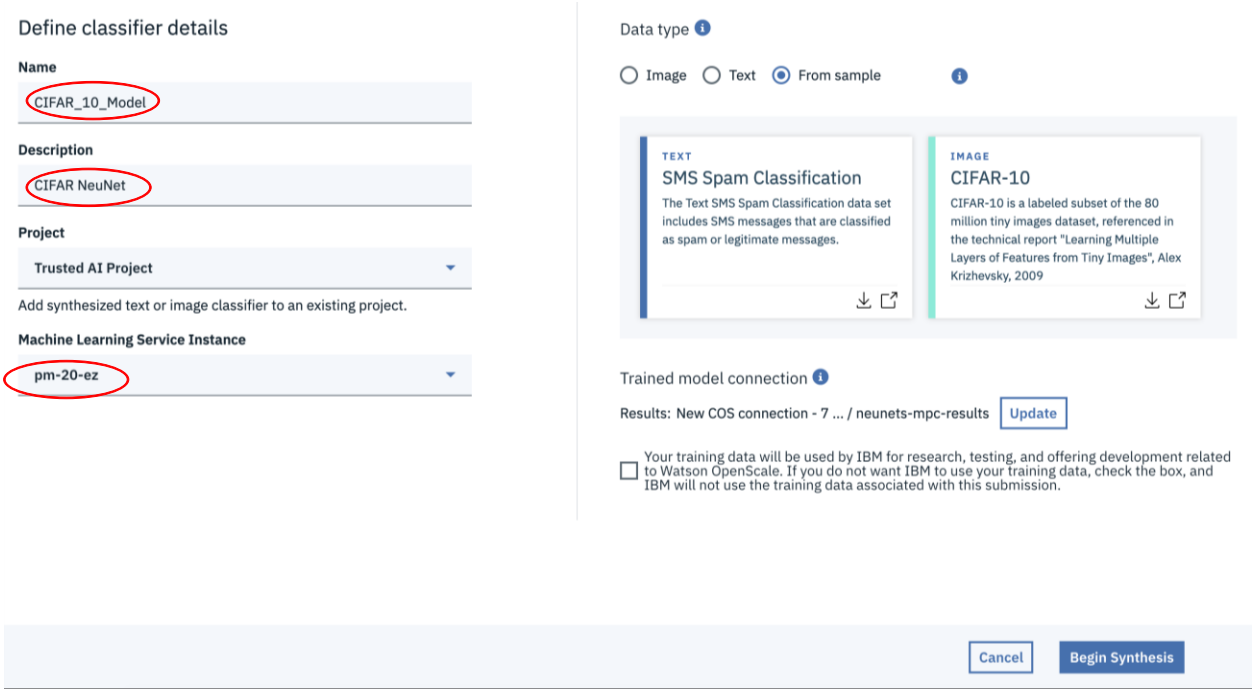
4. Under Trained model connection, click Select.



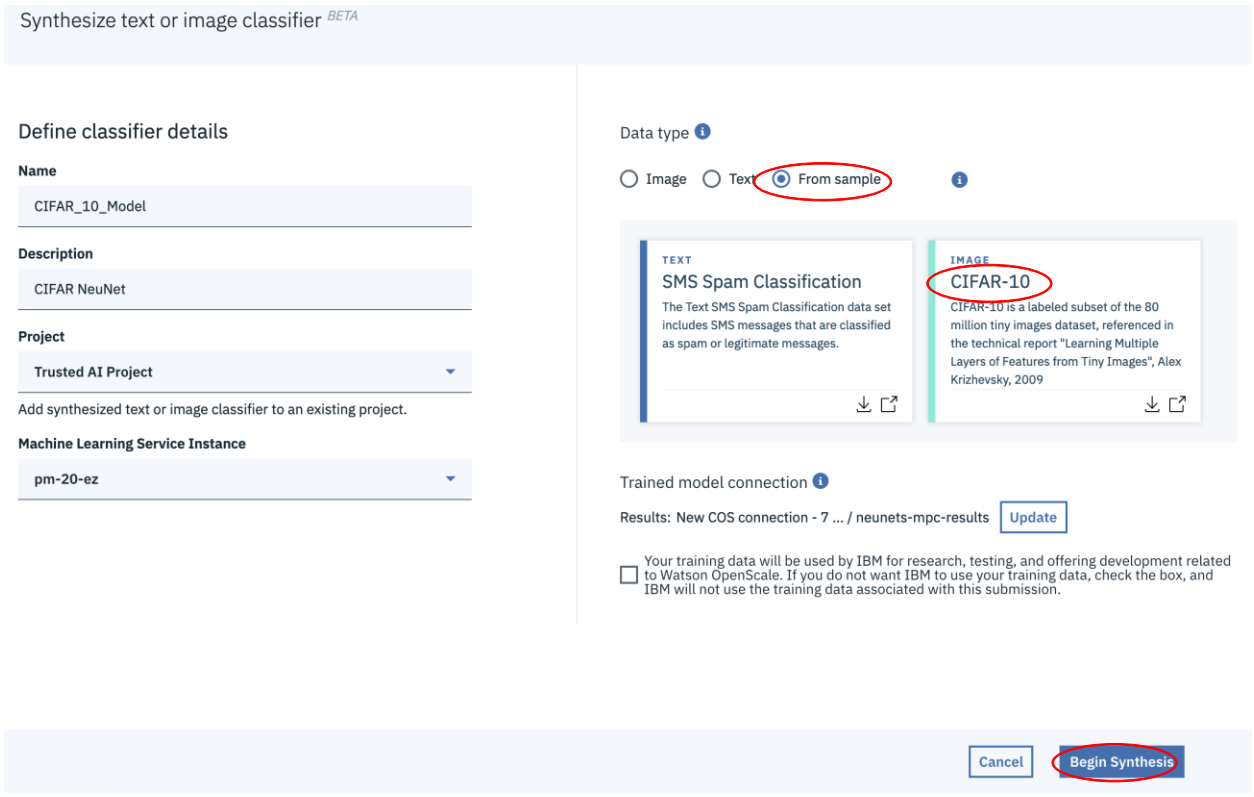
5. Click on the new connection tab.
6. Under Cloud Object Storage instance select your cloud storage form the dropdown.
7. Under Bucket Containing results data enter “neunets-XXXX-results” where the Xs are your initials in lowercase (e.g. neuneuts-mpc-results).
8. Click Create.



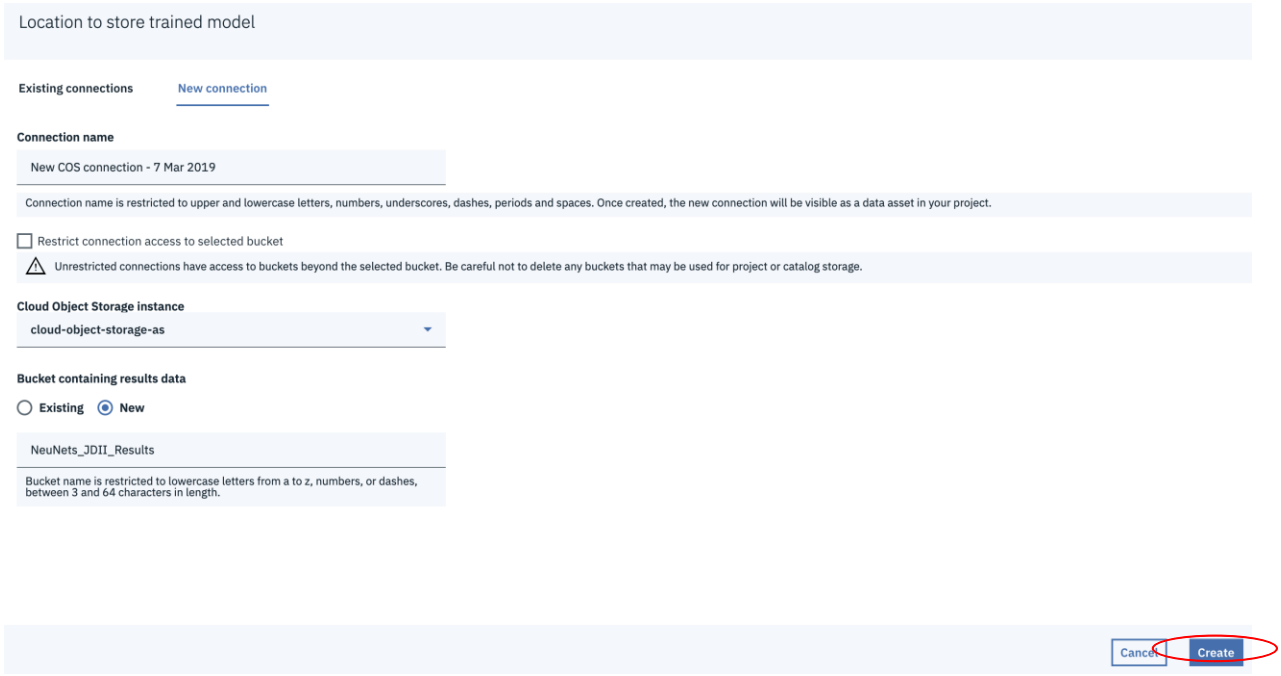
9. Back on the Synthesize text or image classifier page, enter a name and description.
10. Select your Machine Learning service under Machine Learning Service Instance.



11. Select from Sample along the right side and click CIFAR 10.
12. Click Begin Synthesis.

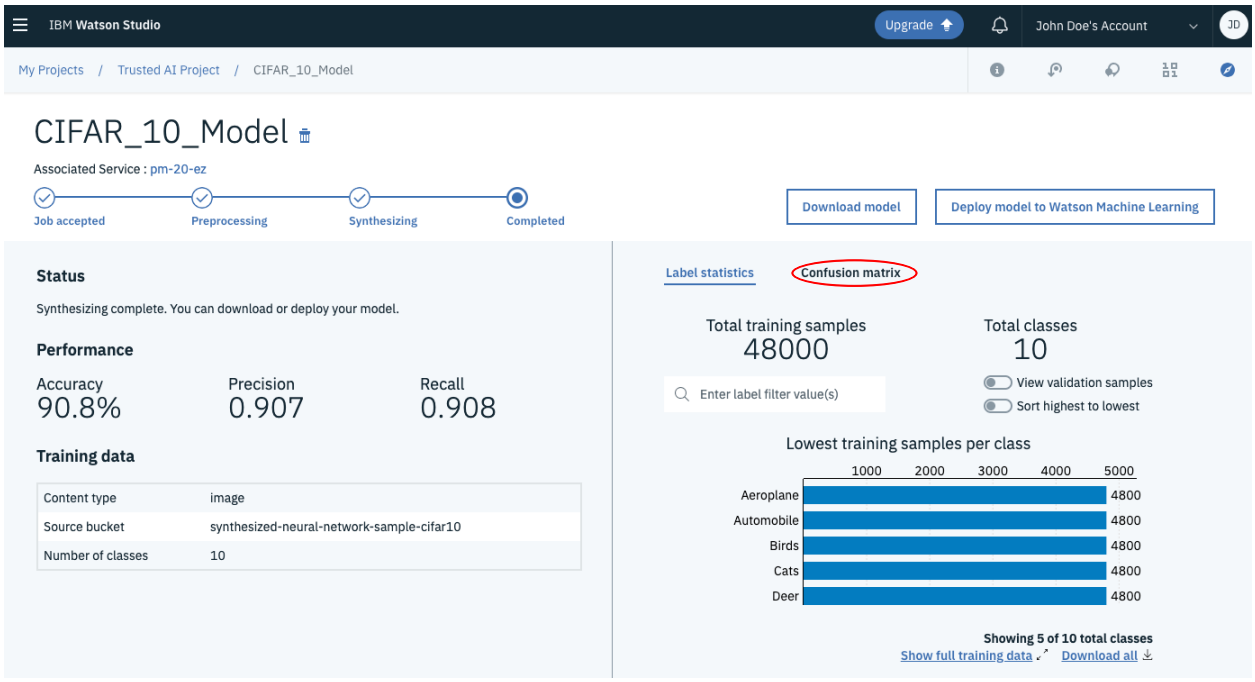


13. Click Create.

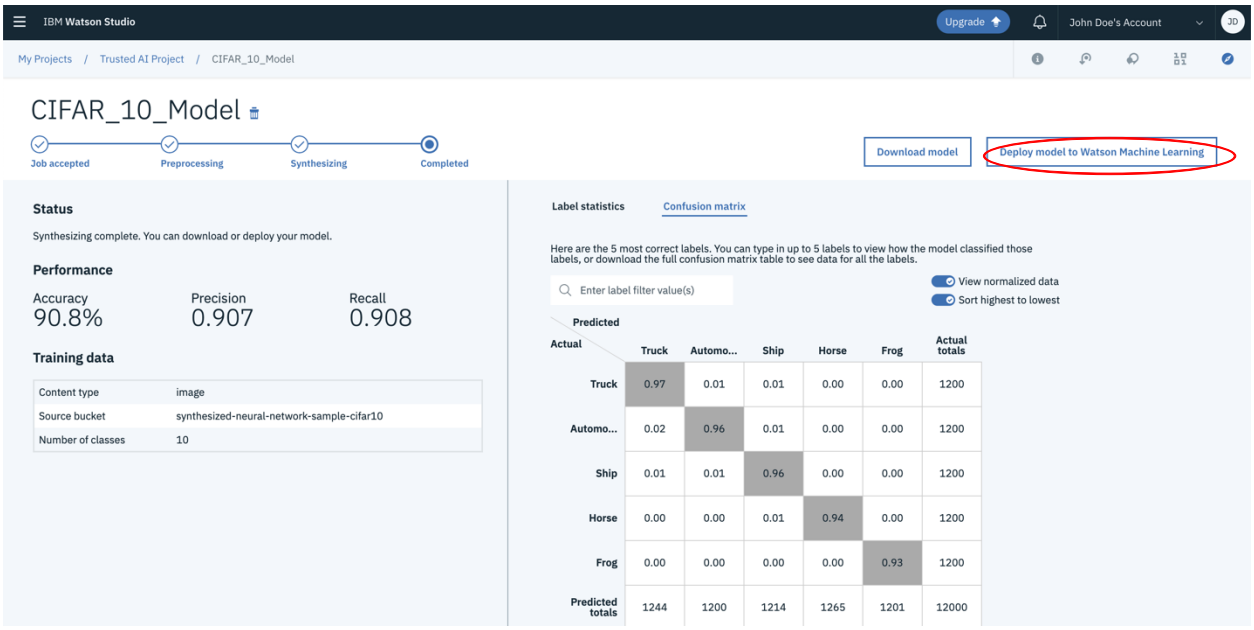


Your model may take a long time to synthesize.

14. Click on Confusion matrix and observe further results.



15. Click on Deploy model to Watson Machine Learning



This is now a Watson Machine Learning Model and can be deployed and monitored using all the techniques shown during the first session today as well as the Watson OpenScale session.

You completed the optional lab.