

# Trial Tutorial for the **H2O AI Hybrid Cloud**



# Welcome!

Welcome to the H2O AI Hybrid Cloud! We are excited to share our platform with you and introduce you to our new trial experience. It has been our mission since its inception to democratize AI. Opening up our cloud experience to the world is an important step in that mission.

To make the best of your 14 day trial, we will walk you through a series of capabilities and applications that showcase what an AI cloud looks like. You will learn how to open, run, and operate AI Apps and even build and deploy your own predictive models. Once you have the basics down, you can explore, experiment, and create freely!

H2O AI Hybrid Cloud is an end-to-end platform that enables organizations to rapidly build world-class AI models and applications in the cloud or on premise. The platform comes with an AppStore that simplifies access to business applications, data science best practices, and allows for rapid

prototyping and collaboration. Built on top of the award winning Driverless AI automatic machine learning platform and the best-in-class H2O open source machine learning library, this platform was designed to accelerate the entire machine learning lifecycle including data ingestion, feature engineering, modeling, interpreting, and deploying models into production. It also enables the rapid development of extremely high quality AI applications covering different end user needs.

In the enterprise version of the H2O AI Hybrid Cloud, you can use our open-source SDK, Wave, to build any style of app you choose. It is a python based SDK that can leverage any python library so you can have complete control and freedom of what apps you develop. We enable the full spectrum of app creation, not just data science. We have provided more than 20 pre-built applications in the trial to give you a sense of what is possible with Wave.

## WHAT TO EXPECT

**The H2O AI Hybrid Cloud is a unified platform composed of several components.**

- **AI application store**
- **AutoML integration**
- **Machine learning interpretability**
- **Automated model documentation**
- **Deployment and model monitoring**

In this trial, we will walk through a use case using supervised machine learning to predict whether or not customers are at risk for churn. Across industries, companies are investing in artificial intelligence to identify customers who are at risk of attrition allowing them to quickly intervene with next best action offers.



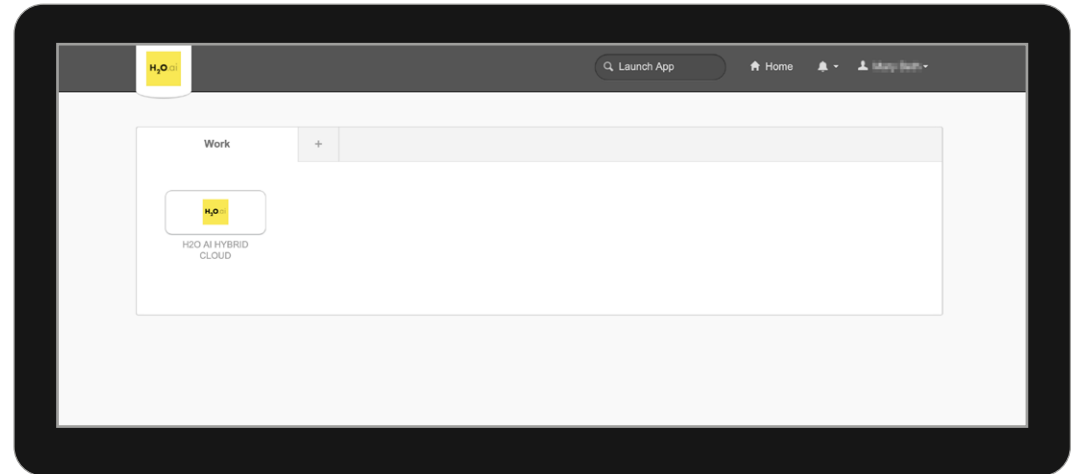
# Explore

You will begin in the App Store which is a central portal for AI applications built using H2O Wave. This includes over twenty sample applications ranging from data science best practice apps to examples of industry vertical solutions. You can also easily provision an AI engine for automatically building predictive models and deploy these models as REST endpoints.

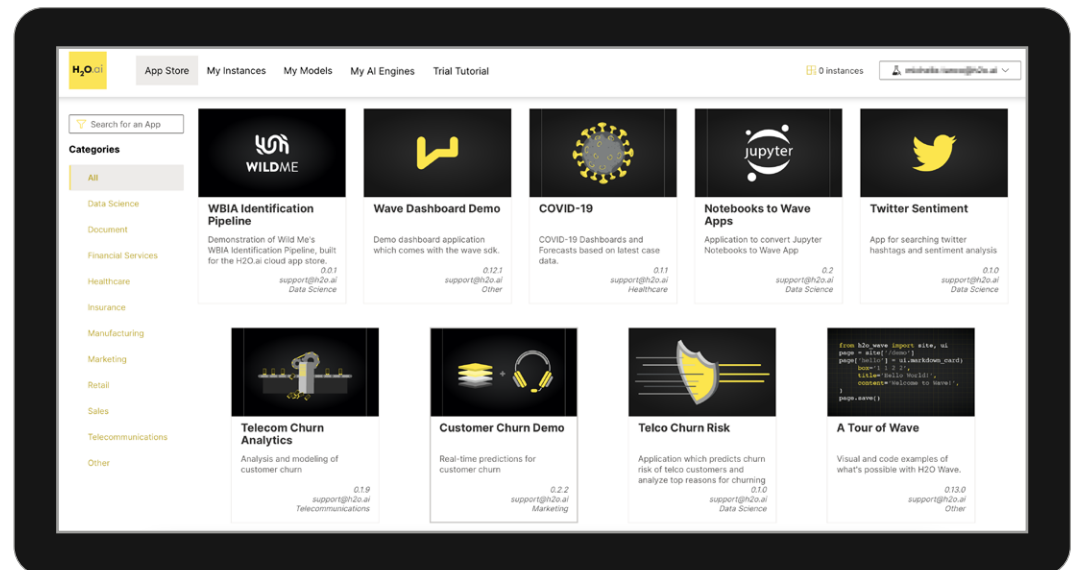
# EXPLORE

## AI APPLICATION STORE

1. Navigate to the **H2O AI Hybrid Cloud** at [cloud.h2o.ai](https://cloud.h2o.ai). You will see a tab labeled **Work** and a button to enter the H2O AI Hybrid Cloud AppStore.

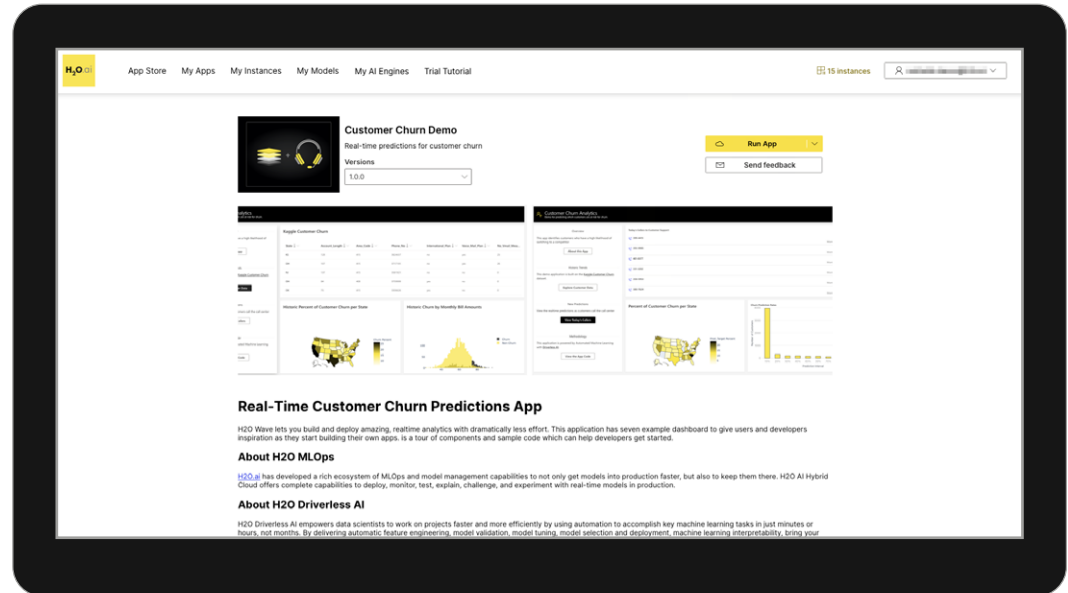


2. In the App Store, scroll through the home page to find the **Customer Churn Demo** application. Let's walk through how a lightweight app rapidly provides insights and predictions on what customers are at risk of churning.



3. On the homepage of an app you can see the following information:

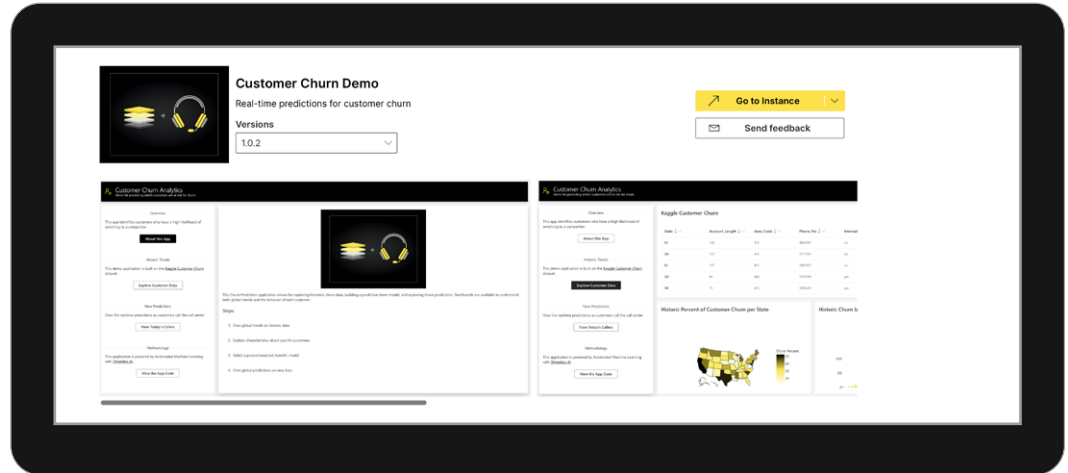
- a. Name and description of the App
- b. A dropdown list of versions
- c. A button to launch a new, separate running instance of the app. If you have an instance of the app already running this button will say **Run App**
- d. Detailed application descriptions from the developer
- e. A list of any running instances of this application



**4. Start an instance of the app by clicking the Run App button**

- a. This will create a new instance of the app in about 45 seconds.
- b. While your instance is being created, scroll to the bottom of the screen to see a list of existing instances. Once yours is ready, you will be able to select ***Pause, Resume*** and ***Terminate*** from the ***Actions*** dropdown menu. You can use the Pause button to shut down the app when not in use to save cloud costs without fully terminating the instance.

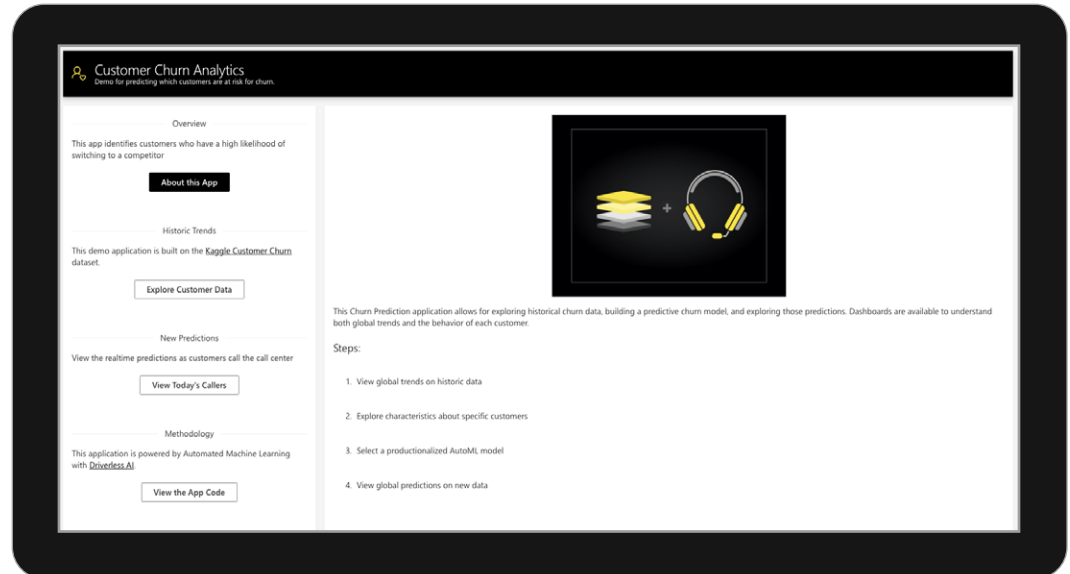
Once complete, the ***Run App*** button on the top right of your screen will change to read **Go To Instance**. Click on the button to run the application.



5. On the **App home screen** you will see information on the purpose of the app, the dataset being used, new predictions and the code behind the app.

6. Click **Explore Customer Data**

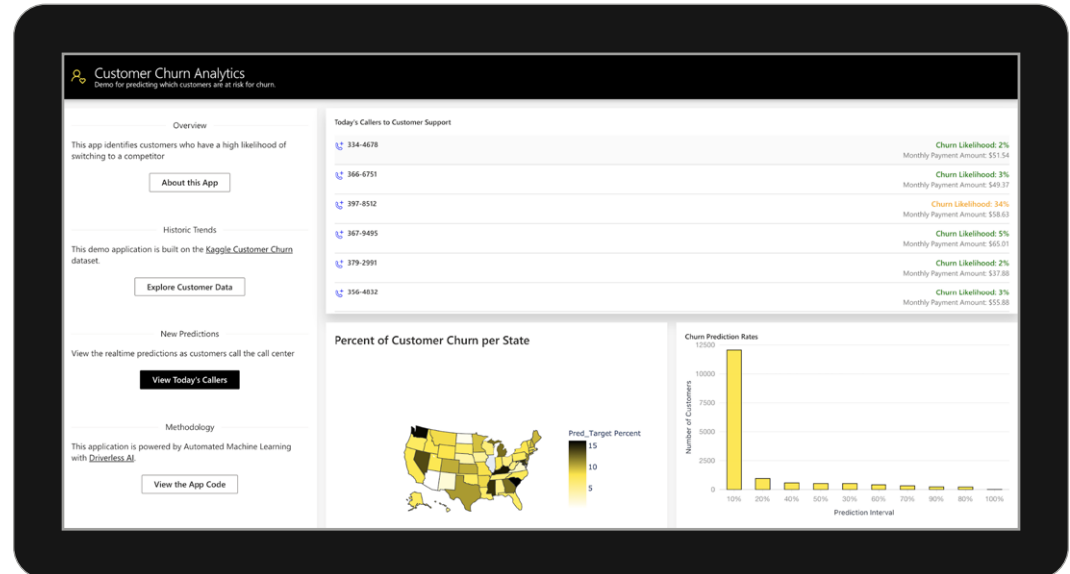
- a. This screen allows us to explore the [Kaggle Telco Customer Churn](#) dataset which has historic information about customers at a telecommunications company and whether or not they have cancelled their contract.
- b. Additionally, we are presented with two visualizations: a geographic analysis and a histogram. We can quickly see that this company is facing severe churn problems in Texas & California as with the accounts with the highest monthly bill amount.





## 7. Click **View Today's Callers**

- a. This page is automatically updated every three seconds with the latest information from our Demo Call Center. As customers call in, their information is scored on a predictive model built using Driverless AI and hosted in MLOps, which we will explore in later steps.
- b. At the top of the page, we can see the most recent callers, their monthly bill amount, and how likely our model thinks they are to churn.
- c. We can also see the current average churn rates by state and a distribution of our model's predictions. We can see that overall the calls coming in we do not see an overall high risk of churn, but certain states, such as Washington and South Carolina, are flagging as higher risk than in the parent model. This could be due to sample size or some data drift factors.



**8. Click View the App Code**

- a. This application was built using the open source framework H2O Wave.
- b. You can explore the Python code that was used to build this frontend application for our customer churn predictive model.

We have successfully walked through our first AI application! In the H2O AI Hybrid Cloud trial, you will have access to 25+ applications. These apps represent just a handful of use cases that the App Store can solve. In the full enterprise version, you can code any application you want and upload to your own personal hybrid cloud.

Close this tab in your browser and return to the **AppStore** tab where you launched your instance.



# Model

Next we will walk through how to create and manage an instance of the AutoML software Driverless AI. For the purposes of this trial users can run one instance of DAI and one instance of H2O-3.





# MODEL

4. You will arrive at a page that allows you to create a new **Driverless AI instance**. There are quite a few configuration parameters available for the user to specify the size and performance capabilities they require for the modeling exercise. Typically, this is a function of the data set size and complexity of the problem the user sets out to solve. For the purpose of this demo, please enter **test-instance** in the text field as shown below and then click **Launch Instance**.

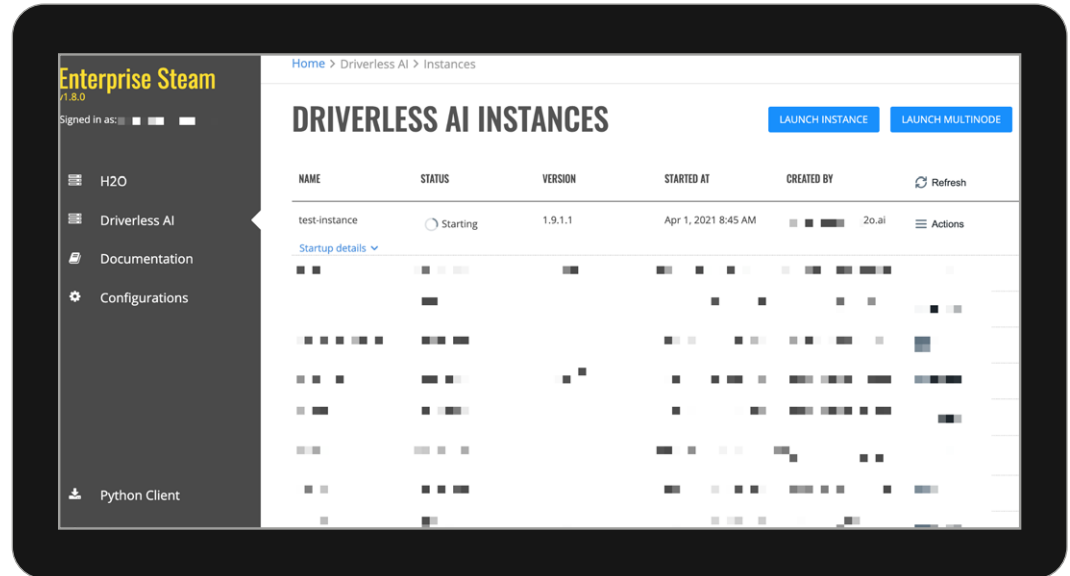
The screenshot displays the 'Enterprise Steam' web interface. On the left is a dark sidebar with the logo 'Enterprise Steam v1.8.2' and the user 'Signed in as: vinod.iyengar@h2o.ai'. The sidebar contains navigation links: H2O, Driverless AI, Documentation, Configurations, Python Client, About, and Support. The main content area shows a breadcrumb trail: Home > Driverless AI > Instances > Launch. Below this is a 'SELECT PROFILE' dropdown set to 'visitor-driverless-kubernet'. A table of configuration parameters follows, with the 'INSTANCE NAME' row highlighted by a green border. The 'INSTANCE NAME' field contains 'test-instance'. Other parameters include VERSION (1.9.0.6), NUMBER OF CPUS (2), NUMBER OF GPUS (0), MEMORY [GB] (12), STORAGE [GB] (20), MAXIMUM IDLE TIME [HRS] (1), MAXIMUM UPTIME [HRS] (1), and TIMEOUT [S] (600). Each parameter row includes a help icon (question mark) and some have additional requirements like 'required min: 1, max: 4'.

Parameter	Value	Requirements	Help
SELECT PROFILE	visitor-driverless-kubernet		
INSTANCE NAME	test-instance	required must meet requirements	?
VERSION	1.9.0.6	required	?
NUMBER OF CPUS	2	required min: 1, max: 4	?
NUMBER OF GPUS	0	required min: 0, max: 0	?
MEMORY [GB]	12	required min: 9, max: 12	?
STORAGE [GB]	20	required min: 20, max: 20	?
MAXIMUM IDLE TIME [HRS]	1	required min: 1, max: 1	?
MAXIMUM UPTIME [HRS]	1	required min: 1, max: 2	?
TIMEOUT [S]	600	required	?

# MODEL

5. After you hit **Launch Instance**, you'll see a screen that brings up the instance, and you can see the progress as the instance is being launched.

Once the instance is created, you can click on the instance name to go directly into **Driverless AI**.



6. Download the **telecom churn data set** from this link:

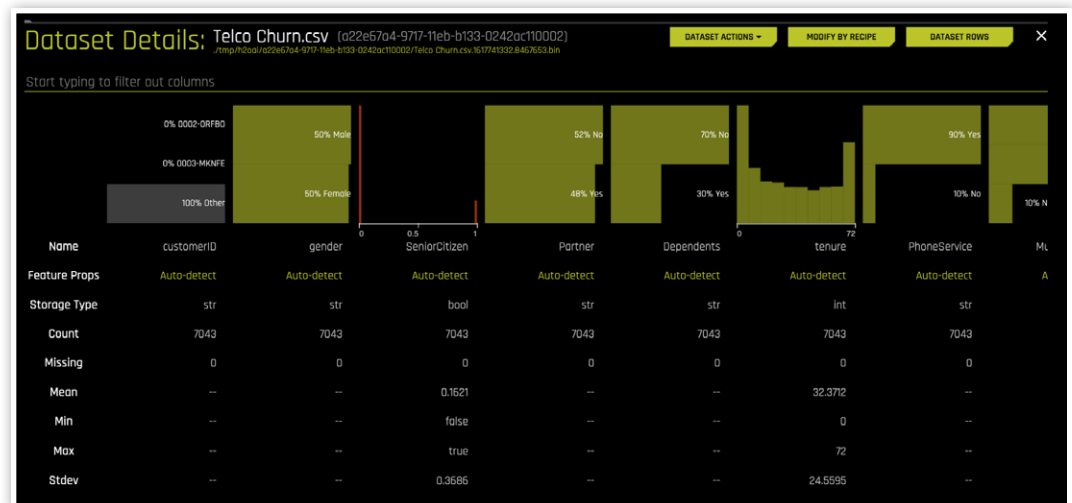
<https://drive.google.com/file/d/1nGEkLFCVd8XGj6LMrvagETeAezaA3WBu/view?usp=sharing>

# MODEL

- Once in Driverless AI, click **Add Dataset**, select upload file and select your **Telco Churn dataset**.



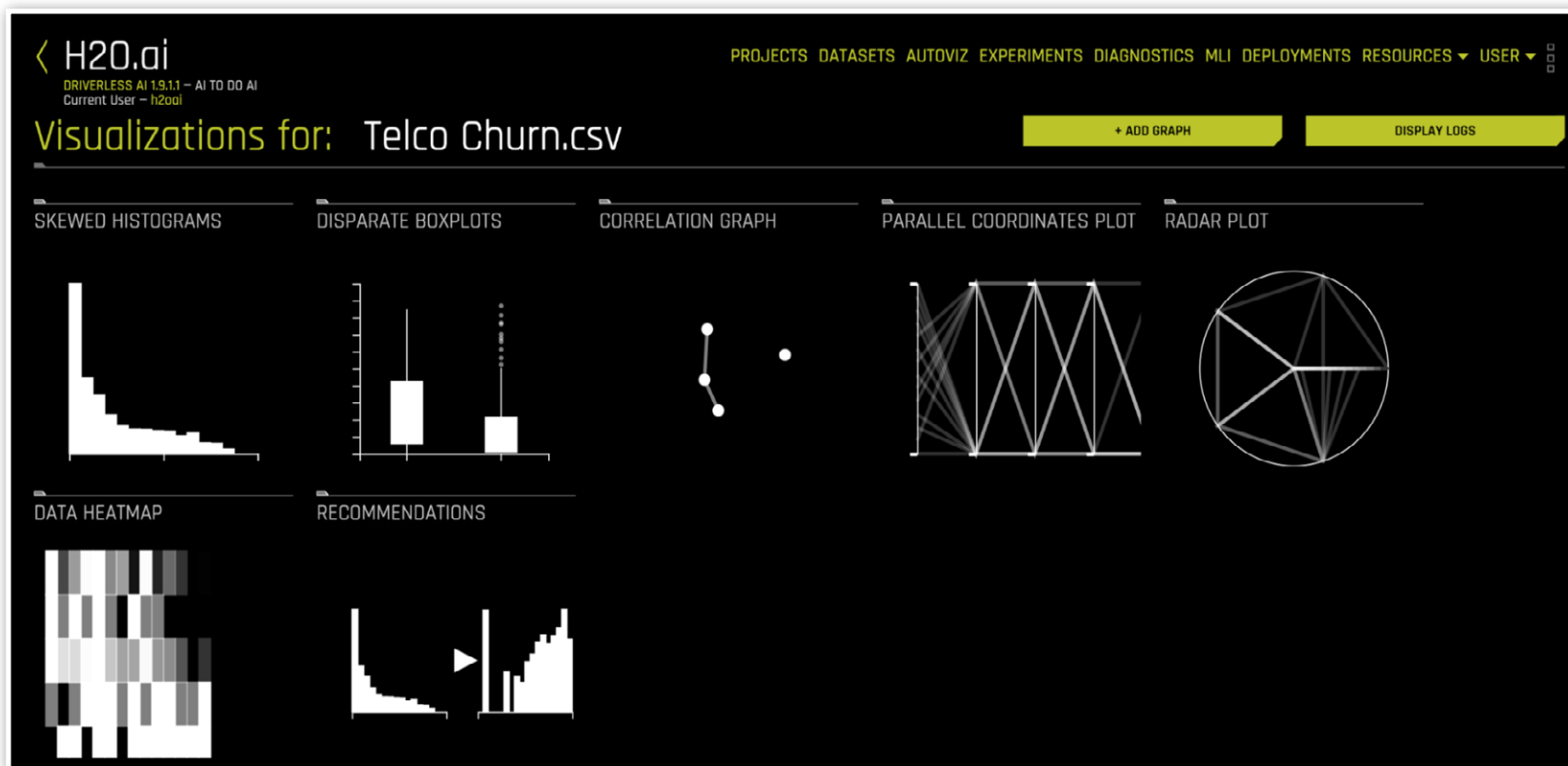
- Let's now explore this data set. Under **Status**, select **Click for Actions** and select **Details** from the dropdown menu. Here we can see (and modify) our data, data types, distributions, and more.





## MODEL

- Click on the yellow x in the top right corner to return to previous screen. Return to the **Click for Actions** menus under Status and this time we will select **Visualize**. Once complete, click on **Visualization ready**, which will appear under the Message column. Here we see **Driverless AI's AutoViz**, an automatically generated series of visualizations that enable users to better understand their data set. Additionally, AutoViz will provide potential recommendations of interesting visuals or data transformations that could help in this experiment. Let's move on to building our model!



10. Click the back arrow beside H2O.ai on the top left of your screen. Under Status, select **Predict** from the **Click for Actions** menu. Set the target column to Churn. We are now presented with a Driverless AI experiment landing page. Here we can adjust the knobs around accuracy, time, and interpretability to tailor our AutoML experiment to our specific problem. Additionally, alternative or custom scoring metrics can be selected or brought in. For more in depth customization of our experiments, we can use the expert settings to enable or disable any capabilities.

**H2O.ai Experiment**  
DRIVERLESS AI 1.9.1.1 - AI TO DO AI  
Current User - h2ool

PROJECTS DATASETS AUTOVIZ EXPERIMENTS DIAGNOSTICS MLI DEPLOYMENTS RESOURCES USER

**EXPERIMENT SETUP** ASSISTANT

DISPLAY NAME: **Telco Churn** TRAINING DATASET: **Telco Churn.csv**

ROWS: **7K** COLUMNS: **21** DROPPED COLUMNS: **--** VALIDATION DATASET: **--** TEST DATASET: **--**

TARGET COLUMN: **Churn** FOLD COLUMN: **--**

WEIGHT COLUMN: **--** TIME COLUMN: **[OFF]**

TYPE: **STR** COUNT: **7043** UNIQUE: **2** TARGET FREQ: **1869**

**TRAINING SETTINGS** EXPERT SETTINGS

ACCURACY: **7** TIME: **1** INTERPRETABILITY: **7** SCORER: **AUC**

CLASSIFICATION REPRODUCIBLE GPUS ENABLED

LAUNCH EXPERIMENT CREATE LEADERBOARD

**What do these settings mean?**

**ACCURACY**

- Training data size: **7,043 rows, 21 cols**
- Feature evolution: **(Constant, DecisionTree, LightGBM, XGBoostGBM), 3-fold CV, 2 reps**
- Final pipeline: **Blend of up to 3 (Constant, DecisionTree, LightGBM, XGBoostGBM) models, each averaged across 3-fold CV splits**

**TIME**

- Feature evolution: **8 individuals, up to 33 iterations**
- Early stopping: **disabled**

**INTERPRETABILITY**

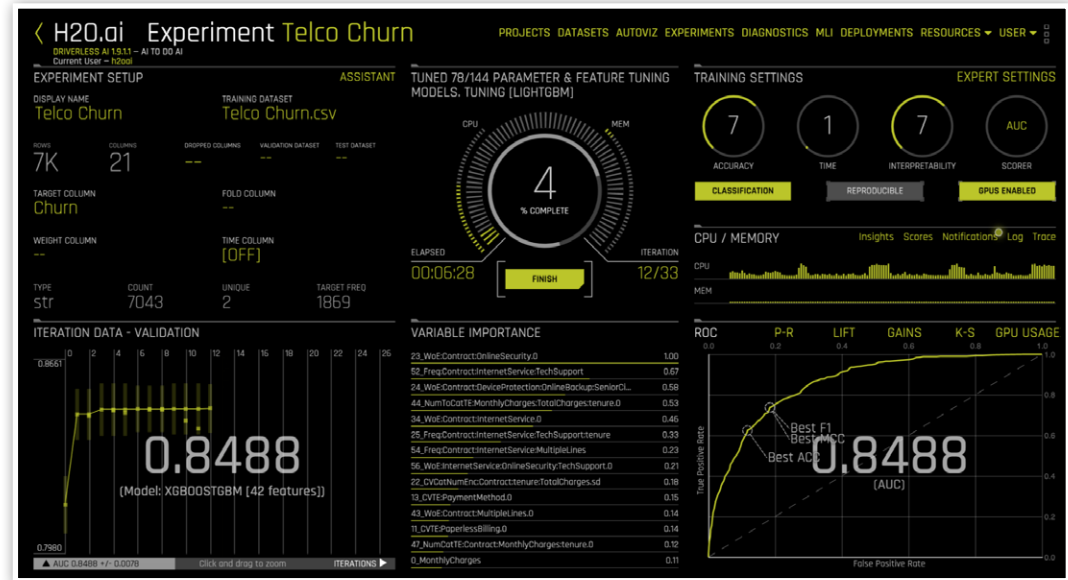
- Feature pre-pruning strategy: **Permutation Importance FS**
- Monotonicity constraints: **enabled**
- Feature engineering search space: **(CVCatNumEncode, CVTargetEncode, CatOriginal, Cat, Frequent, Interactions, NumCatIE, NumToCatIE, NumToCatWoEMonotonic, NumToCatWoE, OneHotEncoding, Original, TextLinModel, TextOriginal, Text, WeightOfEvidence)**
- Pre-trained PyTorch NLP models (with fine-tuning): **['disabled']**

[Constant, DecisionTree, LightGBM, XGBoostGBM] models to train:

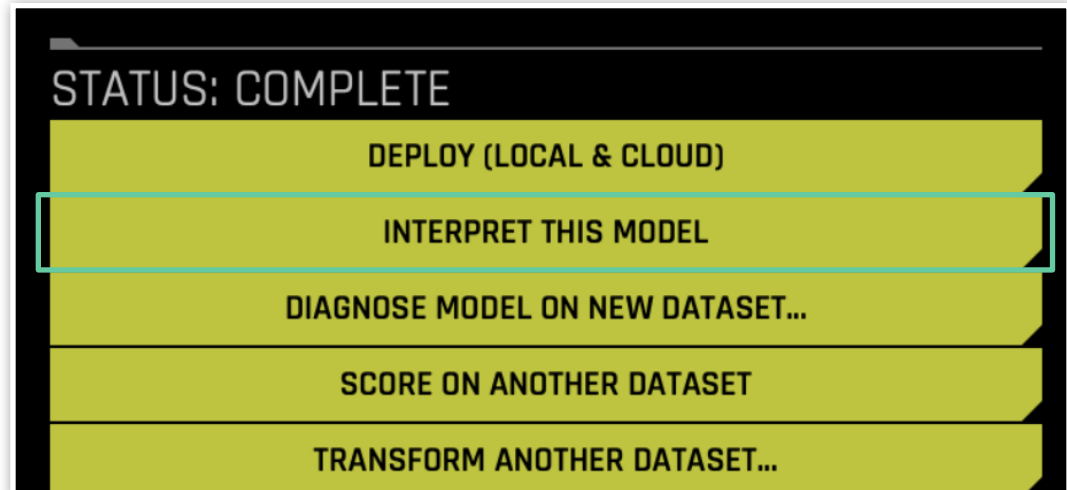
- Model and feature tuning: **192**
- Feature evolution: **72**
- Final pipeline: **9**
- Per-model Hyper. opt. trials: **0 (evolution) 0 (final)**

Estimated runtime: **minutes**  
Auto-click Finish/Abort if not done in: **1 day/7 days**

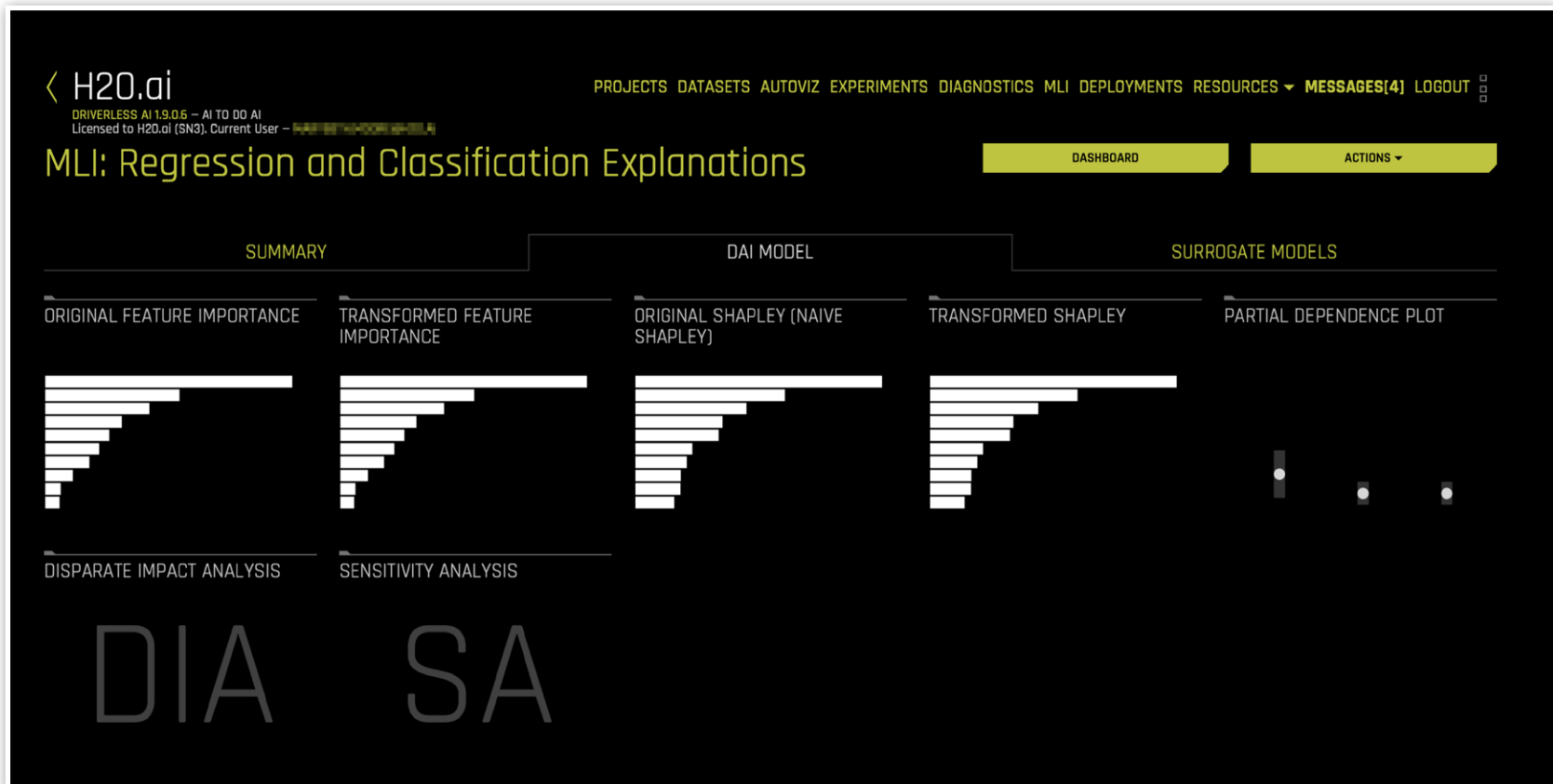
**11.** Click **Launch Experiment**. You can watch as the AutoML builds and tests many models and features to determine what the best champion model will be.



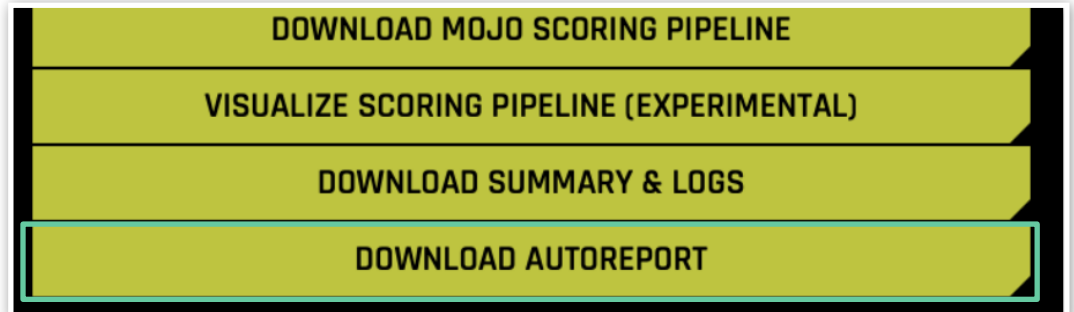
**12.** Once the model is complete, there will be several options presented for what to do next. Let's select **Interpret this Model**.



- 13.** Here we can see a variety of explainer capabilities and visualizations to help you understand what your model is thinking. Here you can check for bias/fairness, run what-if scenarios, download your model as a rules-based system, and more! Click the yellow back arrow at the top left of your screen to return to the last page.



14. We will now look at our model documentation that has been automatically created for us. Driverless AI will automatically generate a model document that outlines the entire breadth and depth of the process the system went through to develop the model, the features it created, and the tests it ran. Click on **Download Autoreport** and explore your model documentation.



15. With this brief introduction, you can start running experiments of your own! Driverless AI has been built to enable data scientists to solve their most complex problems with AutoML. For more info on all of the customization options and deeper technical insight, check out our documentation here:

<http://docs.h2o.ai/driverless-ai/latest-stable/docs/userguide/index.html>

**-Driverless AI Experiment: Telco Churn**  
Generated by: h2oai  
Generated on: Tue Apr 6 21:05:22 2021

Driverless AI Experiment: Telco Churn ..... 1  
 Experiment Overview ..... 1  
 Data Overview ..... 1  
 Methodology ..... 1  
 Data Sampling ..... 1  
 Validation Strategy ..... 2  
 Model Tuning ..... 2  
 Feature Evolution ..... 2  
 Feature Transformations ..... 2  
 Final Model ..... 2  
 Alternative Models ..... 2  
 Deployment ..... 2  
 Partial Dependence Plots ..... 2  
 Shapley Values ..... 2  
 Appendix ..... 3

**Experiment Overview**

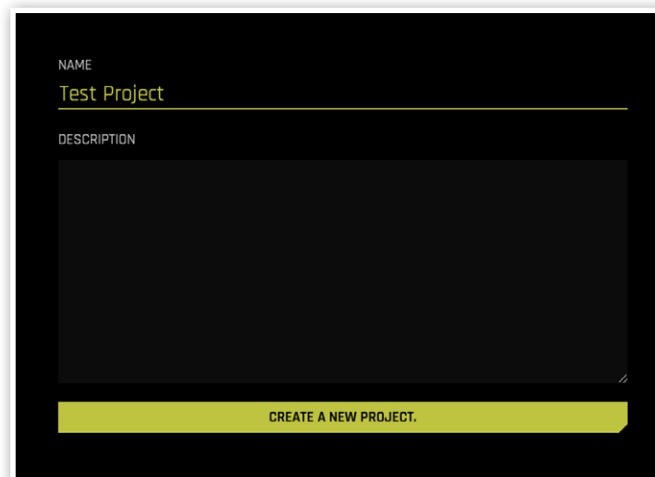
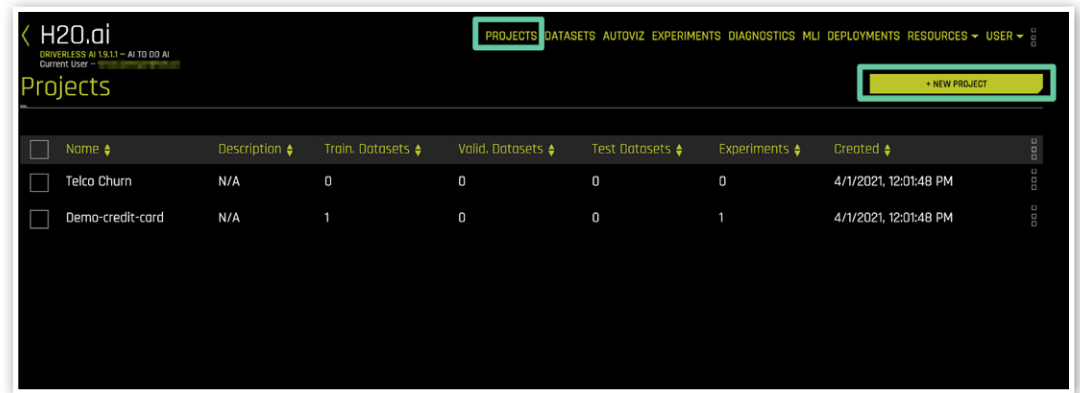
Driverless AI built a stacked ensemble of 1 XGBoostGBMModel, 1 LightGBMModel to predict *Churn* given 20 original features from the input dataset *Telco Churn.csv*. This classification experiment completed in 19 minutes and 20 seconds (0:19:20), using 4 of the 20 original features, and 38 of the 41 engineered features.

**Performance**

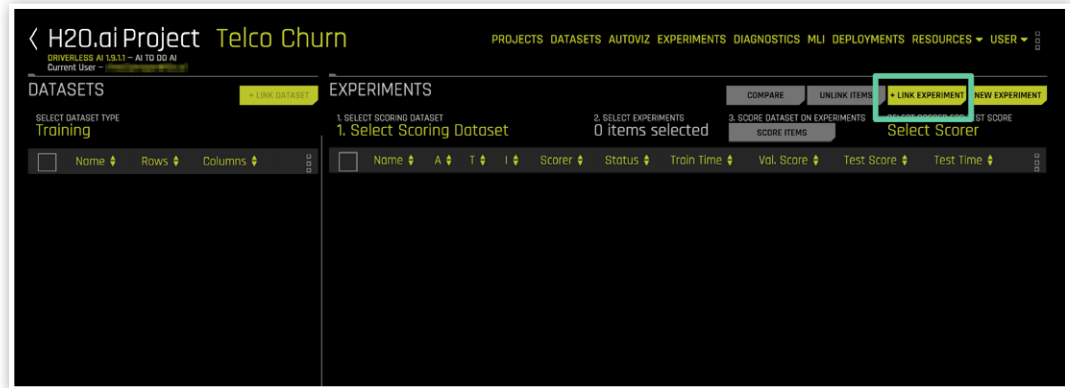
Dataset	AUC
Internal Validation	0.85
Test Data	Test Data not Provided

# Deploy

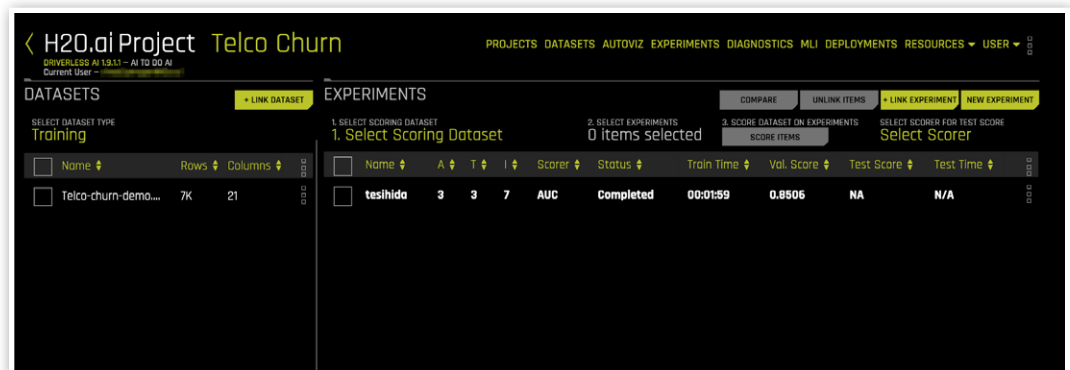
1. To get started with deployment, let's navigate to the **Projects** tab at the top of the screen.



2. Click on **Link Experiment** on the sub-navbar and select the experiment we just completed earlier.



You'll now see the experiment show up on the screen.

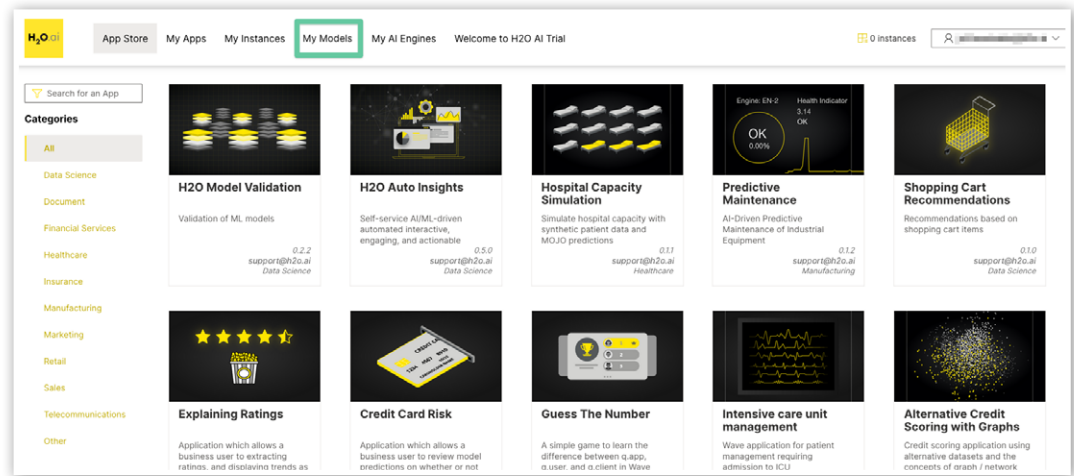




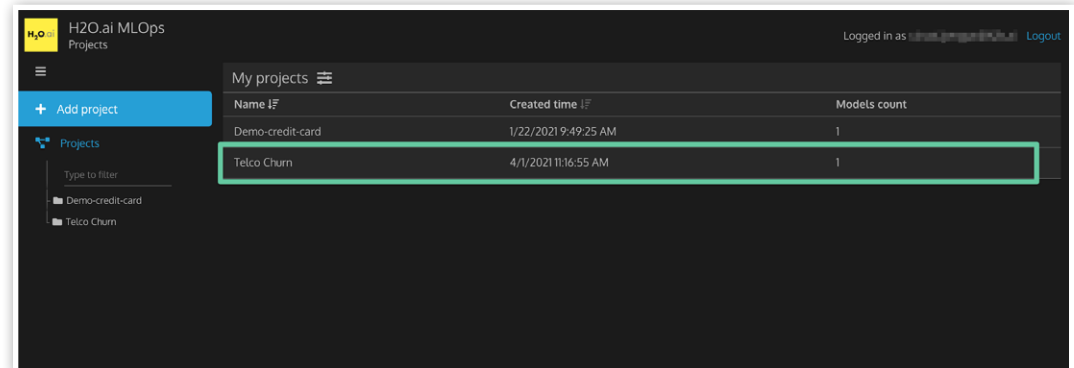
# DEPLOY

## DEPLOYMENT AND MODEL MONITORING

- Return to the H2O AppStore, which is opened as a separate tab in your browser. In the AppStore, click on the **My Models** tab.



You can now see the project show up in MLOps.



# DEPLOY

## DEPLOYMENT AND MODEL MONITORING

- Click on the project name and you can see the model show up there. Click on the **Actions** menu on the right hand side.

The screenshot shows the H2O AI interface for the 'Telco Churn' project. The 'Summary' section displays project details: Name (Telco Churn), Project ID (68017...2b03f), Created on (4/1/2021 11:16:55 AM), Models (1), Deployments (0), Alerts (0, 0, 0), and Owner (vinod.iyengar@h2o.ai). The 'Deployments' section shows 'No deployments available.' The 'Models' table lists the following model:

Model name	Model ID	Created on	Model source	Scorer	Score	Owner	Environment	Tag	Actions
tesihida	b38cd...2573e	4/1/2021 12:...	DAI	AUC	0.85064144...	vinod.iyeng...	No deploym...		...

The 'Actions' menu for the 'tesihida' model is open, showing options: Deploy to DEV, Deploy to PROD, Challenger for ..., Download Autoreport, and More details.

You are now ready to **deploy** the model!



# Discover

H2O-3 is a fully open source, distributed in-memory machine learning platform with linear scalability. H2O-3 supports the most widely used statistical and machine learning algorithms, including gradient boosted machines, generalized linear models, deep learning, and more. H2O.ai also has an industry leading AutoML functionality that automatically runs through all the algorithms and their hyperparameters to produce a leaderboard of the best models. H2O-3 is used by over 18,000 organizations globally and is extremely popular in both the R and Python communities.

Learn more about H2O open source or walkthrough our tutorial here:

<https://training.h2o.ai/products/1a-introduction-to-machine-learning-with-h2o-3-classification>

# Next Steps

You have now familiarized yourself with the H2O AI Hybrid cloud and its capabilities.  
If you want to learn more about how H2O.ai can help you solve your most  
challenging problems with AI schedule a demo by emailing us at:

[sales@h2o.ai](mailto:sales@h2o.ai)

