**AMAZON MACHINE LEARNING**

**What is Amazon Machine Learning?**

Amazon Machine Learning (Amazon ML) is a robust, cloud-based service that makes it easy for developers of all skill levels to use machine learning technology. Amazon ML provides visualization tools and wizards that guide you through the process of creating machine learning (ML) models without having to learn complex ML algorithms and technology. Once your models are ready, Amazon ML makes it easy to obtain predictions for your application using simple APIs, without having to implement custom prediction generation code, or manage any infrastructure.

**Amazon Machine Learning Key Concepts**

This section summarizes the following key concepts and describes in greater detail how they are used within Amazon ML:

* **Datasources** contain metadata associated with data inputs to Amazon ML
* **ML models** generate predictions using the patterns extracted from the input data
* **Evaluations** measure the quality of ML models
* **Batch predictions asynchronously generate** predictions for multiple input data observations
* **Real-time predictions synchronously generate** predictions for individual data observations

**Regions and Endpoints**

Amazon Machine Learning (Amazon ML) has two endpoints that support HTTPS requests. You can use these endpoints for the full range of Amazon ML functionality.

|  |  |  |  |
| --- | --- | --- | --- |
| **Region name** | **Region** | **Endpoint** | **Protocol** |
| US East (N. Virginia) | us-east-1 | machinelearning.us-east-1.amazonaws.com | HTTPS |
| EU (Ireland) | eu-west-1 | machinelearning.eu-west-1.amazonaws.com | HTTPS |

**Solving Business Problems with Amazon Machine Learning**

We can use supervised ML approaches for these specific machine learning tasks:

- Binary classification (predicting one of two possible outcomes),

- Multiclass classification (predicting one of more than two outcomes) and

- Regression (predicting a numeric value).

**Using Amazon ML to Predict Interest Rate for FreddieMac Single-Family-Housing-Loan**

Step 1: Prepare Your Data

Step 2: Create a Training Datasource

Step 3: Create an ML Model

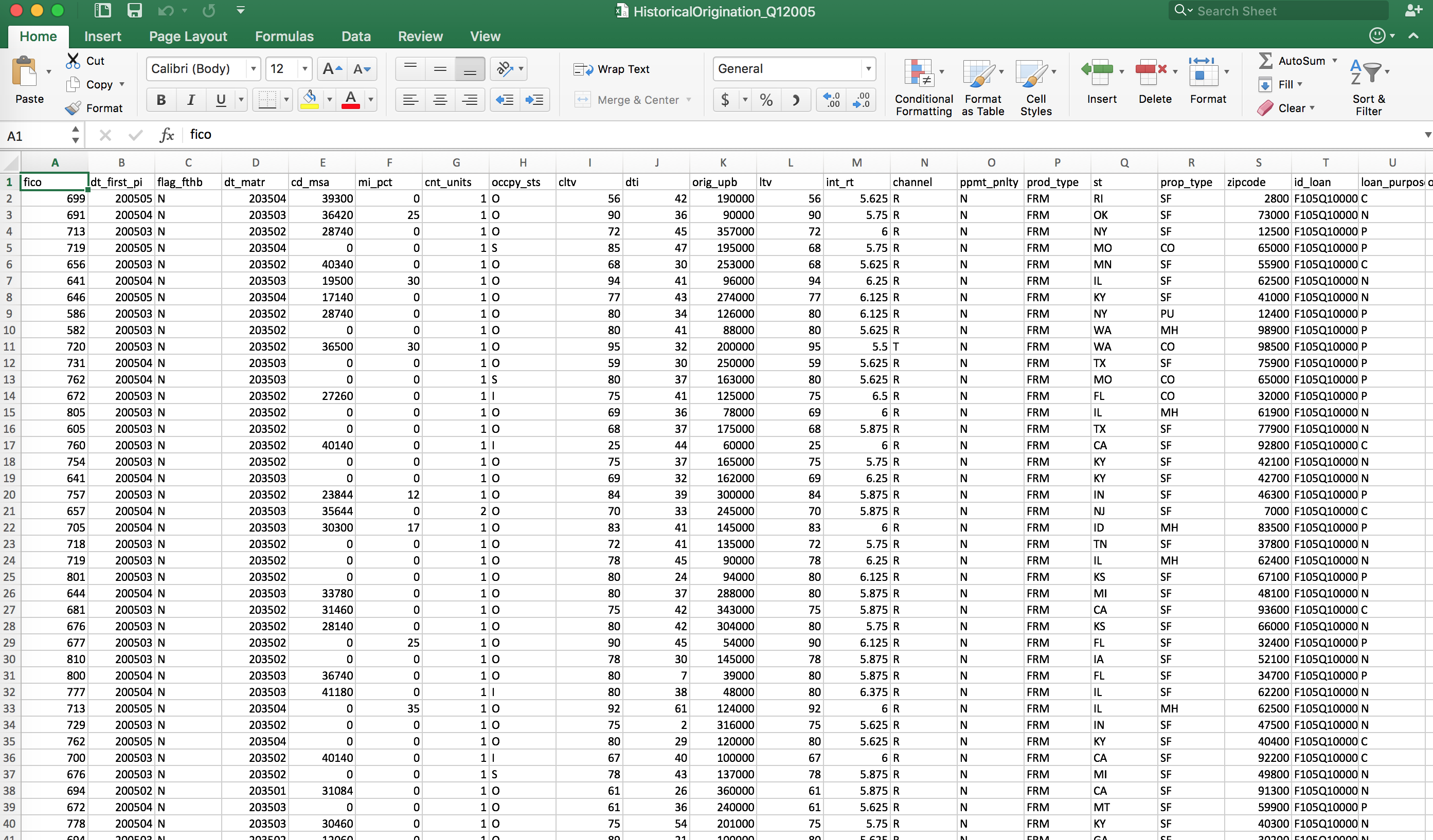
Step 4: Review the ML Model's Predictive Performance and Set a Score Threshold

Step 5: Use the ML Model to Generate Predictions

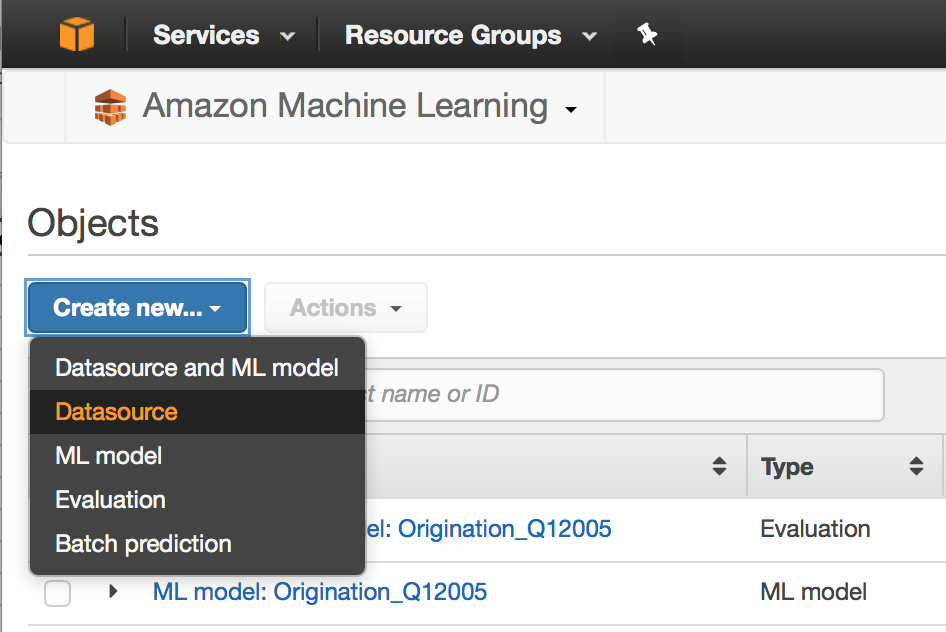
**Step 1: Prepare Your Data**

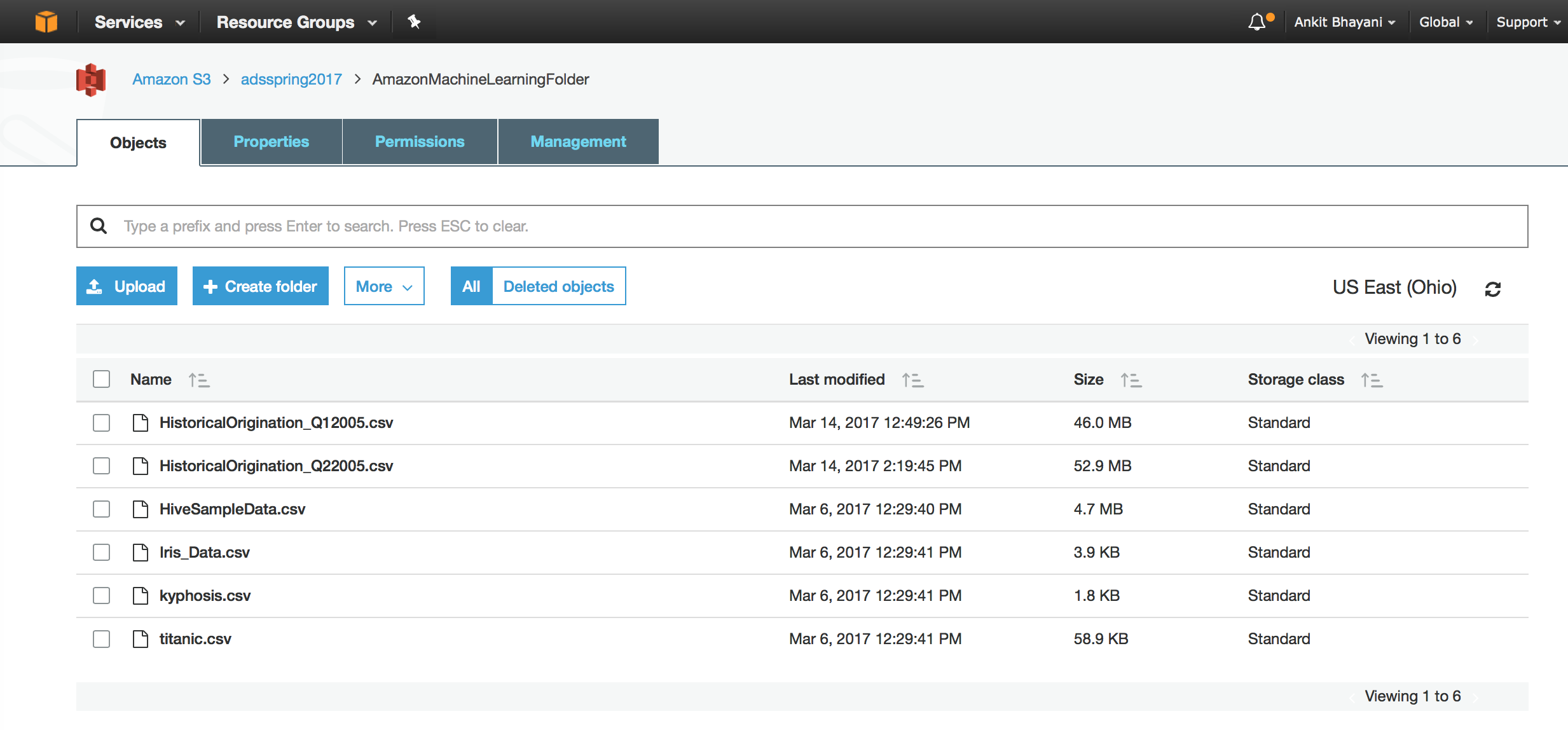
- Preprocess the raw input data.

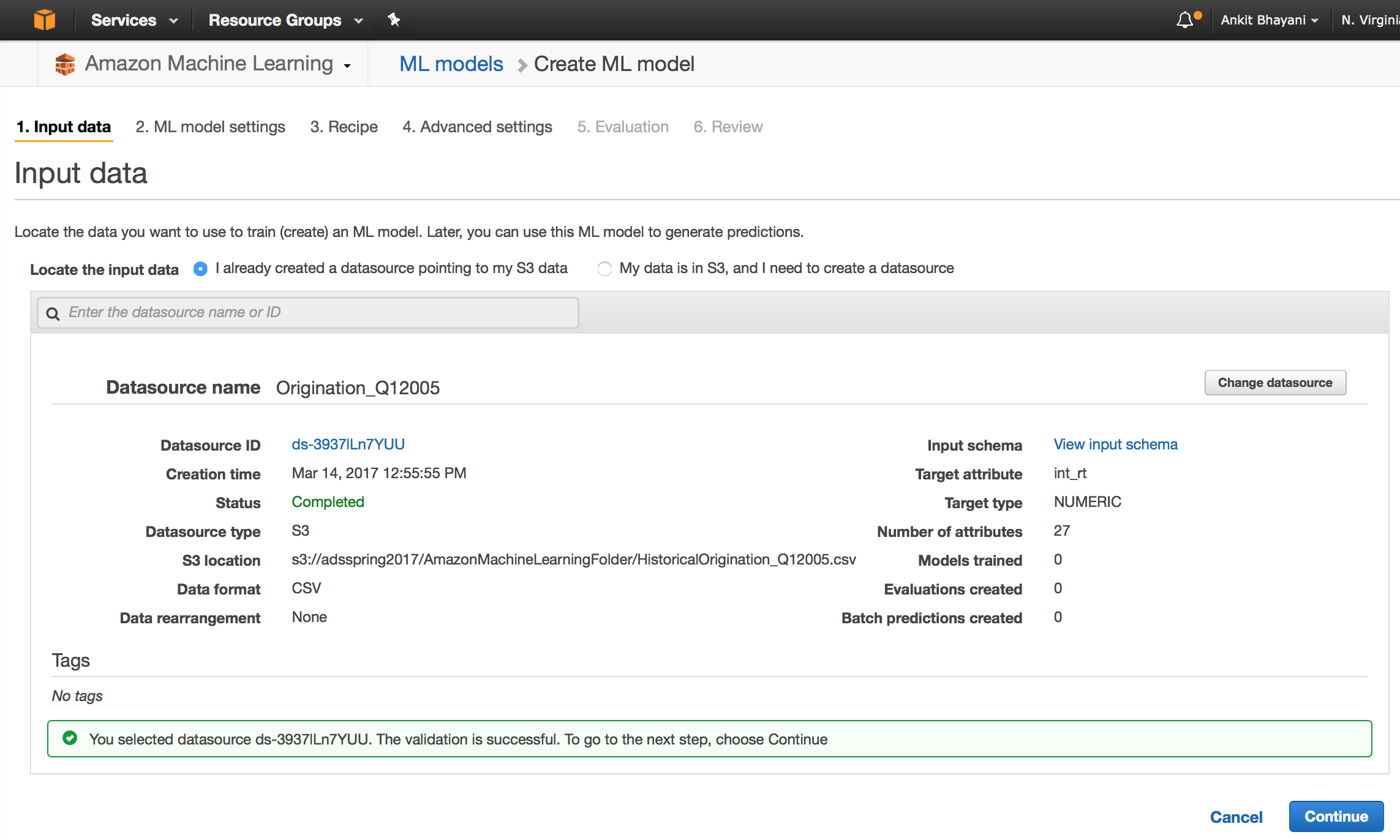
- Remove Nan, fill the missing value, derived columns

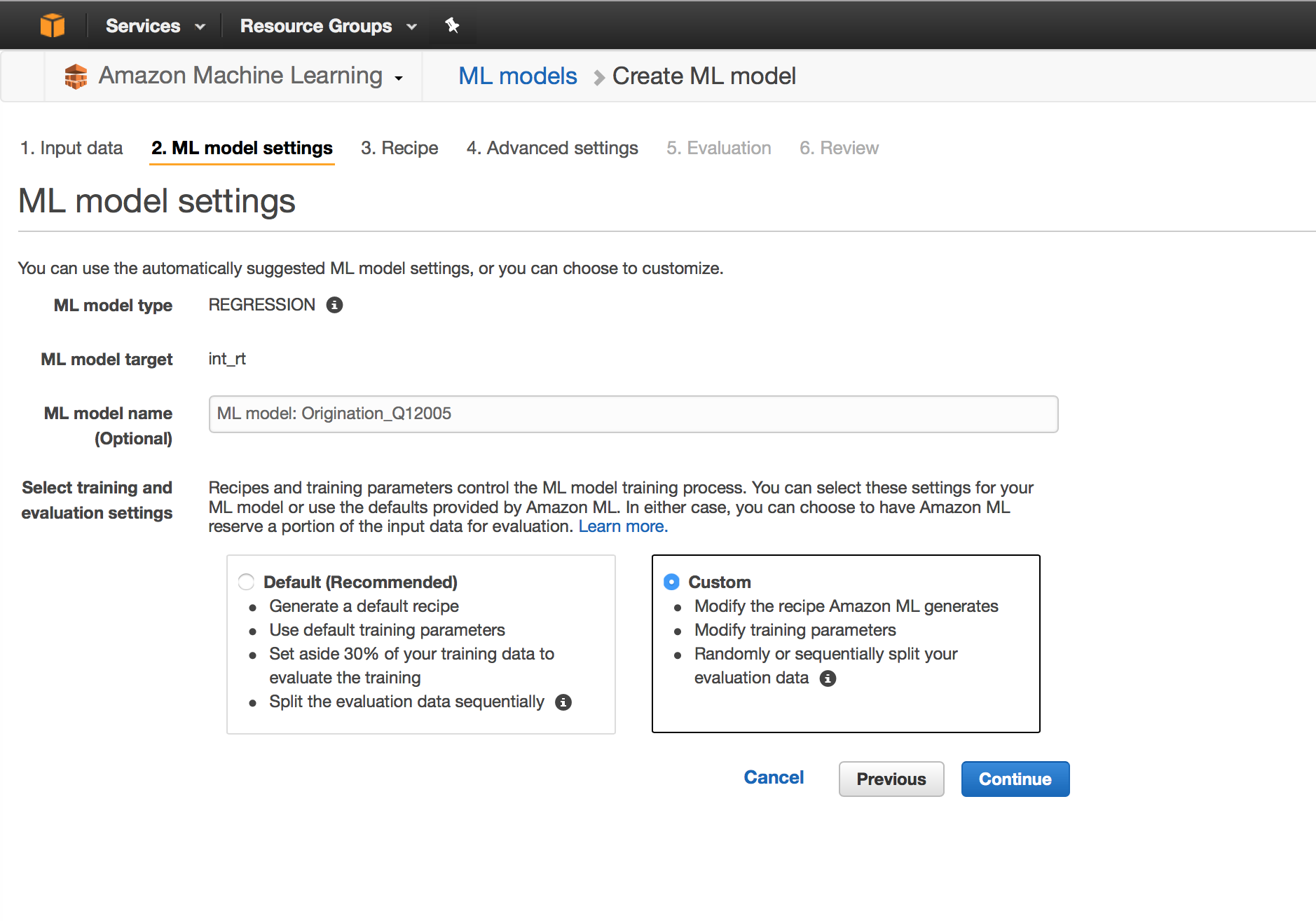
****

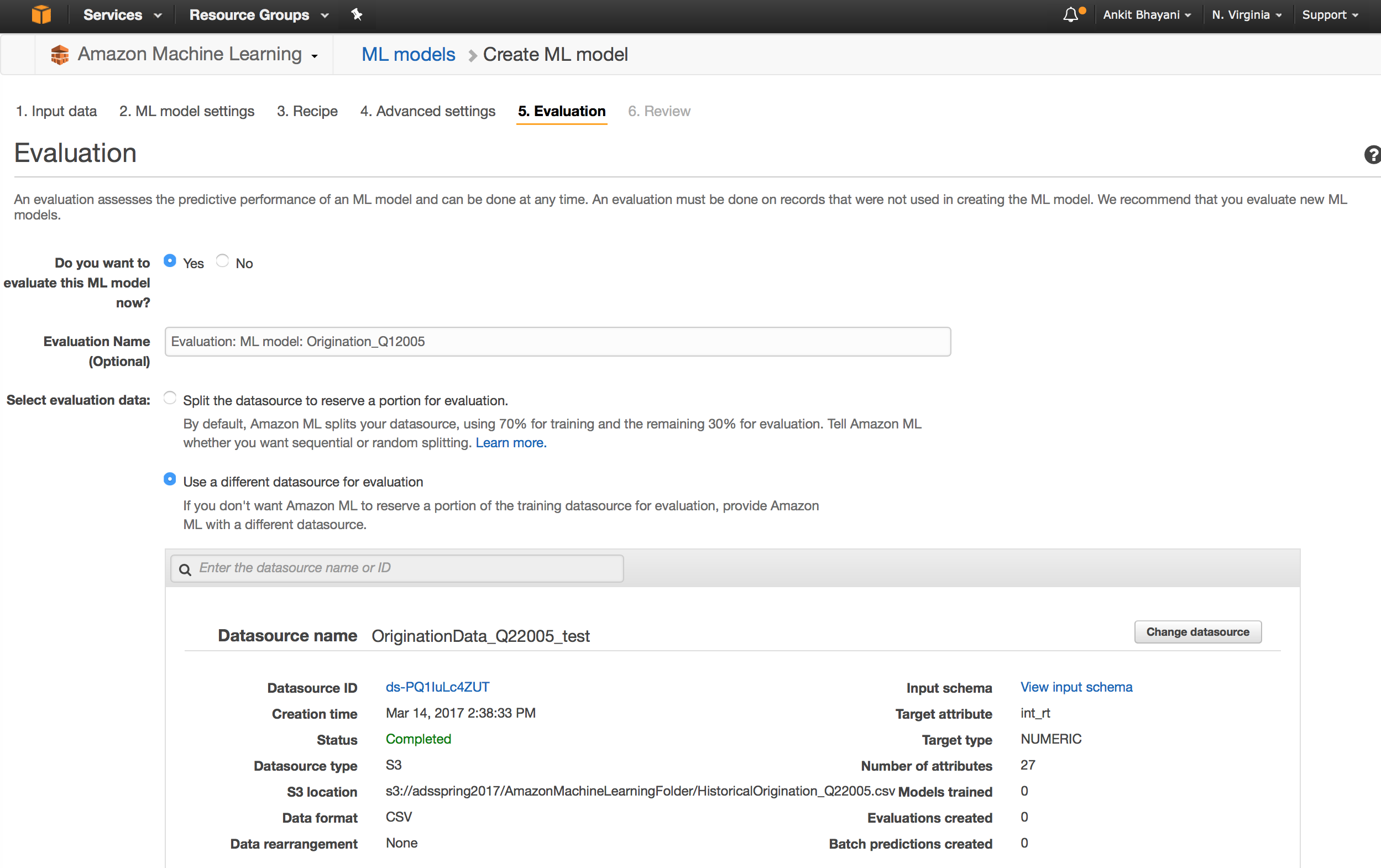
**Step 2: Create a Training and Test Datasource**

****

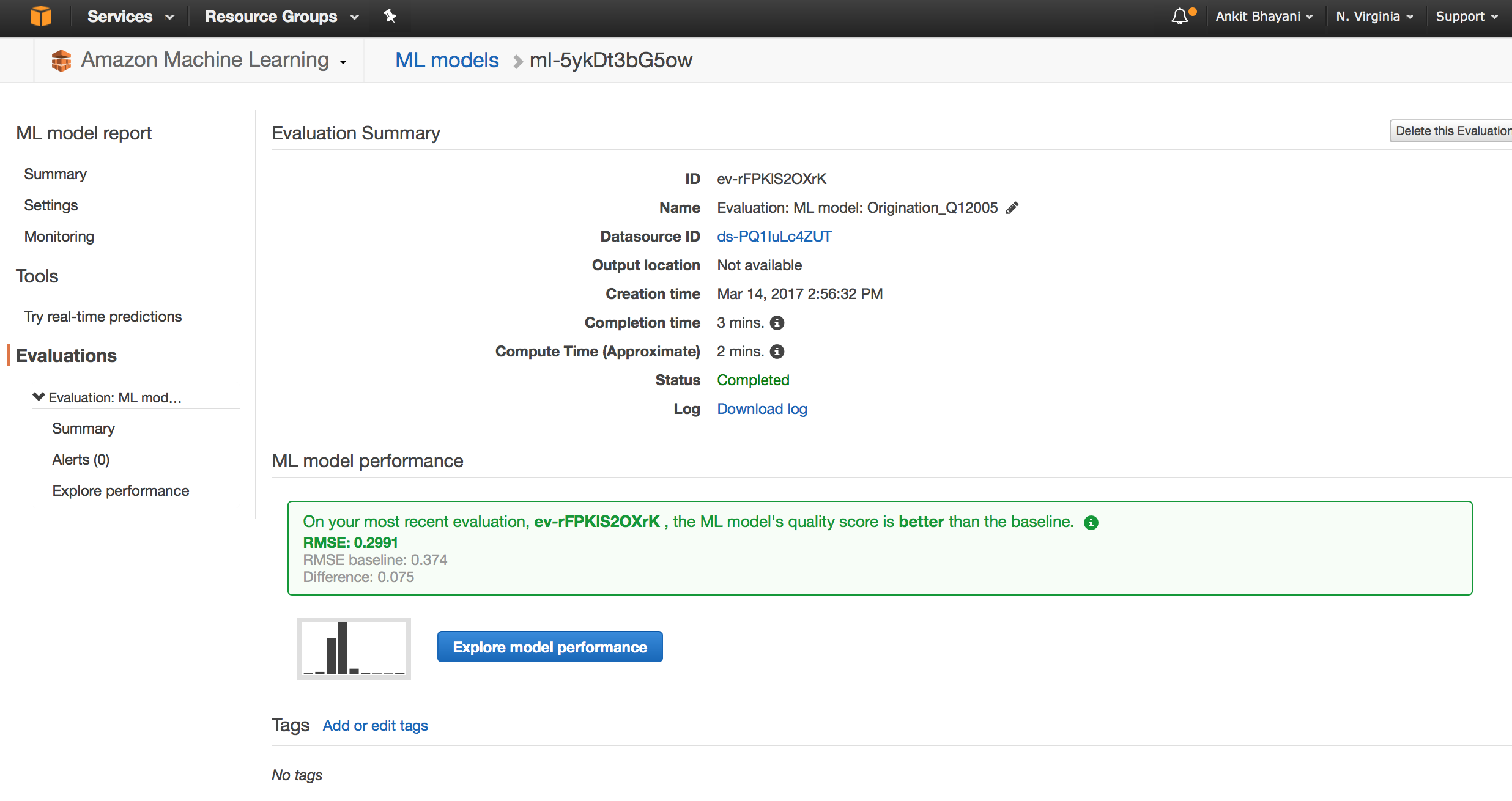
****

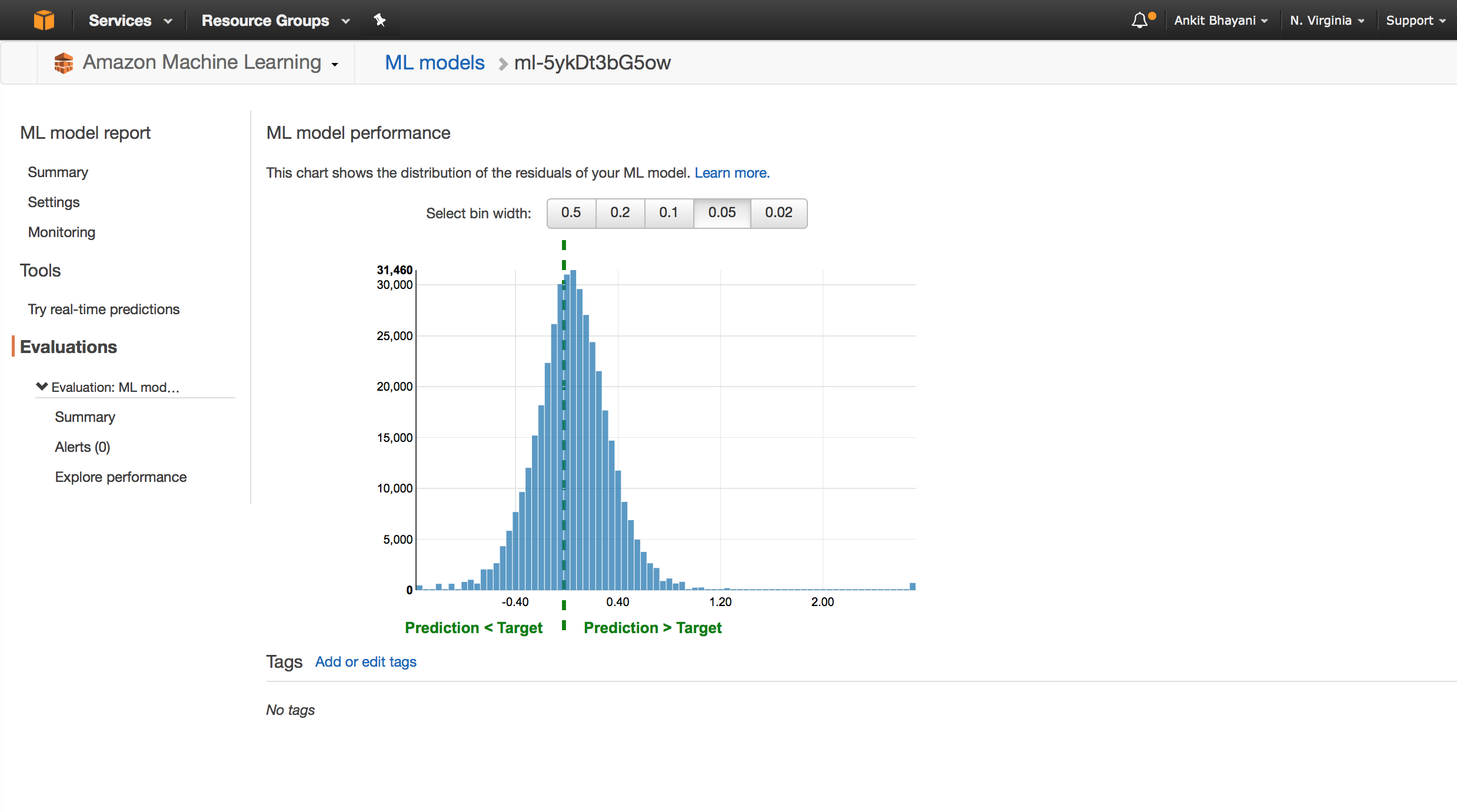
**Step 3: Create an ML Mode**

****

****

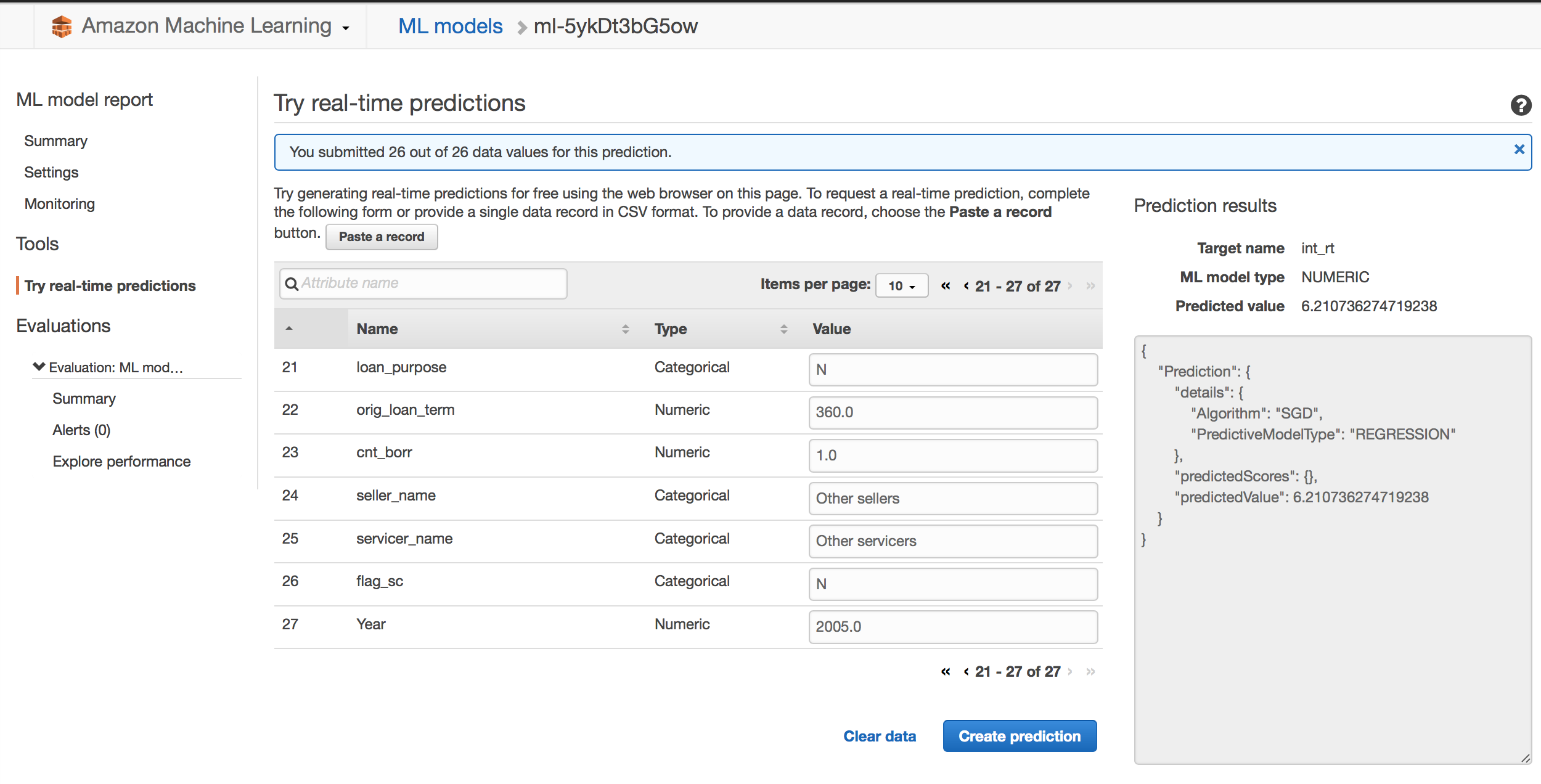
**Step 4: Review the ML Model's Predictive Performance and Set a Score Threshold**

****

****

**Step 5: Use the ML Model to Generate Predictions**

Expected Value: 6.25

Output from AWS: 6.21  ****

**Performance wise**:

* Amazon’s Machine Learning (AML) clearly produced a better result than my best model created from Scikit-Learn’s.
* AML is also extremely easy to use - It takes more days to implement of my Scikit-Learn’s models, yet with AML, total time taken was less than 30 minutes.

**Price wise**:

- Creating a model cost almost nothing with AML (except for the time taken, which is quite long ~ 5 min, almost 5x comparing to Scikit-Learn’s model!).

- However, keep in mind that it’s quite expensive to run your prediction, according to [Amazon Machine Learning’s Pricing](http://aws.amazon.com/machine-learning/pricing/) it cost $0.10 per 1000 batch predictions and $0.0001 per real-time prediction.

**Practical wise**:

* AML gives us a production-ready service, scalability and deployment should be easily done. This is, in my opinion, the best-selling point of AML.
* In some cases, AML is not very flexibly configurable to your special or domain oriented business needs, but most of the case, I strongly think that it should give a very good baseline approach to solve business’s data science problem effectively.

References:

<http://docs.aws.amazon.com/machine-learning/latest/dg/tutorial.html?icmpid=docs_machinelearning_console>