INDIANA UNIVERSITY
BLOOMINGTON

ENGR-E516 Engineering Cloud Computing

Mid Term Report

Jayesh Prasad Anandan

Afeefa Bano

Navya Sree Santhapeta

March 27th, 2023

Retail Sales Prediction and Application

using AWS SageMaker, AWS Lambda and AWS Quicksight

Project Goals

To develop a machine learning based application that predicts the retail sales of an enterprise with the help of XGBoost Algorithm implemented using AWS Sagemaker. Later, to deploy the endpoint of this model as an API (Application Programming Interface) with the help of AWS Lambda and AWS API Gateway. Finally, to create a user-friendly webpage to allow users to input data and generate sales predictions.

Related Work

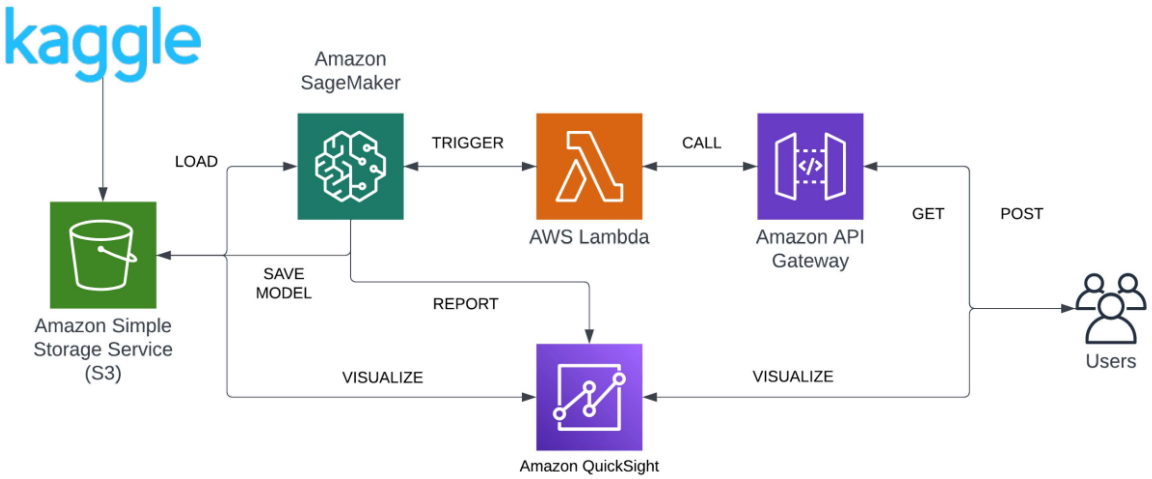
There have been several related works in the field of retail sales prediction using machine learning techniques. Some of the notable ones are as follows:

1. "Predicting Future Sales and Profit Margin of Retail Stores using Machine Learning" by Vijayabhaskar et al. (2018): In this study, the authors proposed a predictive model for retail sales and profit margin using machine learning techniques. The model was trained on a dataset of historical sales data and used a combination of regression and clustering algorithms to generate accurate forecasts.
2. "Retail Sales Forecasting using Machine Learning Algorithms" by Srinivas et al. (2020): In this study, the authors proposed a retail sales forecasting model using machine learning algorithms, including XGBoost. The model was trained on a dataset of historical sales data and used several input features, including product attributes, promotions, and seasonal trends, to generate accurate sales forecasts.
3. "Time Series - ARIMA, DNN, XGBoost Comparison" from Kaggle Blog Post. This blog is about comparison of three models; ARIMA, DNN and XGBoost on Time Series Dataset. The format of the data set on which the comparisons were drawn matches the format of the data set on which we can make our own predictions.

All these related works highlight the importance of accurate sales forecasting in retail and the potential benefits that can be gained from using machine learning techniques. The proposed project, Retail Sales Prediction Using AWS SageMaker XGBoost (Regression), aims to build upon these existing works by using AWS SageMaker's managed training and hosting platform to create a scalable and accurate predictive model for retail sales forecasting.

Proposed Method

We will kick start the project by configuring AWS environment and grant access to all team members. Initially, we will upload our dataset of retail sales prediction to our AWS S3 bucket for further use. Next, we will carry out data preprocessing and exploratory data analysis (EDA). Further, the prediction model will be implemented in AWS Sagemaker and trained with the retail sales data. Once the model is trained, we will be testing the model with untrained data to observe the model's accuracy. Finally, we will implement hyperparameter tuning for finding the optimalhyperparameters for the model.



For the second half of the project, we will use AWS Lambda to get an endpoint to the AWS Sagemaker and connect it with AWS API Gateway to create an API for our model. We can use this API to create a website where a user can utilize our model to find future sales predictions. Also, we will be creating a dashboard from our data to visualize the trends, and insights into the sales. Finally, we will consolidate our observations and prepare a comprehensive report.

Tasks

Week 1: System environment configuration and Dataset upload in AWS S3. - Completed

Week 2: Preprocessing and EDA on the Data using AWS SageMaker. - Completed

Week 3: Implement Prediction Models on the data in AWS SageMaker. - Completed

Week 4: Hyper-parameter tuning and Integration of AWS Lambda. - Completed

Week 5: Creation of API Gateway for our model. - Completed

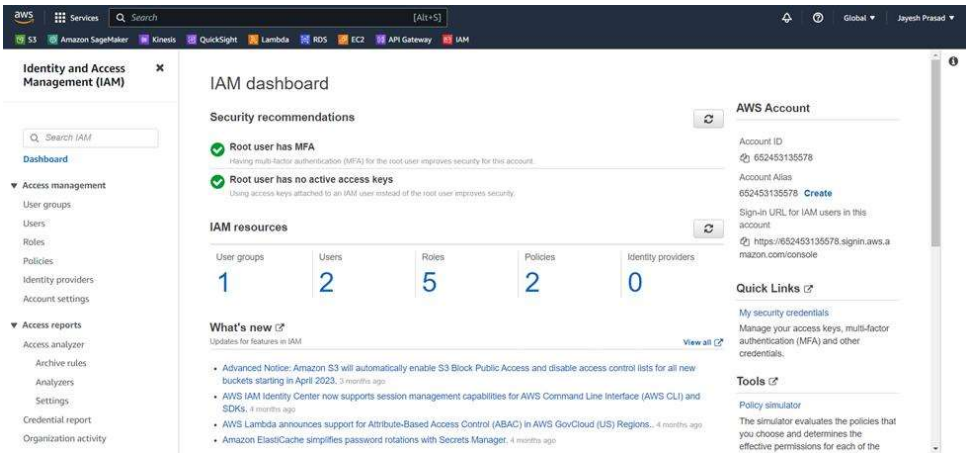
Week 6: Development of Website to utilize the Model. - In Progress

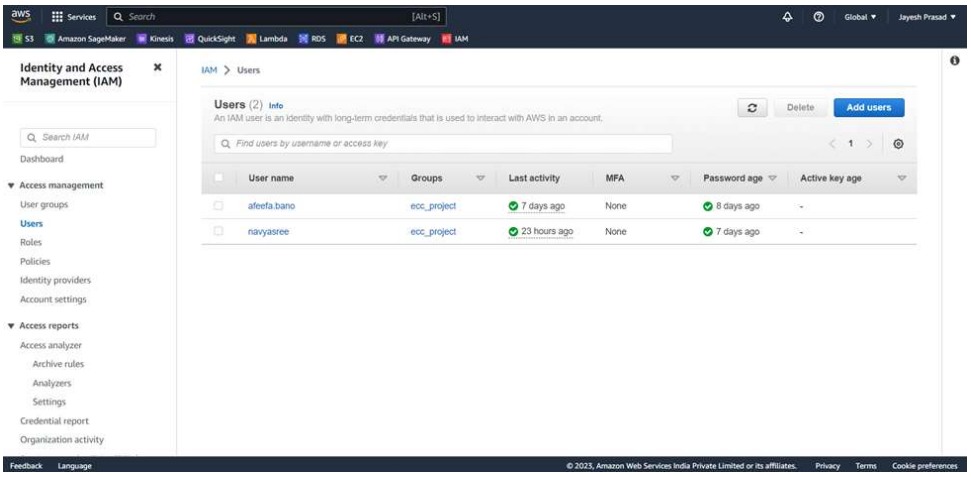
Week 7 Creating Dashboard to visualize the data using AWS Quicksight. - Yet to Start

Week 8: Summarize observations and submit the project report. - In Progress

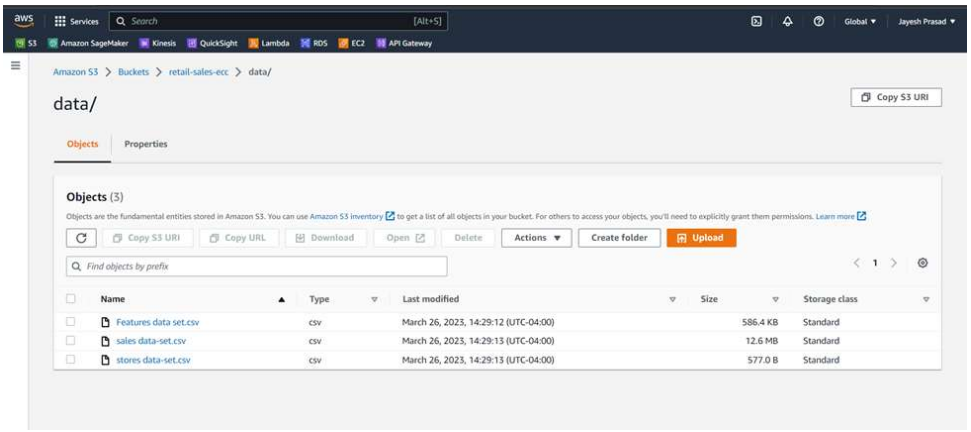
Progress

1. Successfully configured AWS Environment and added teammates to the account as IAM Users.

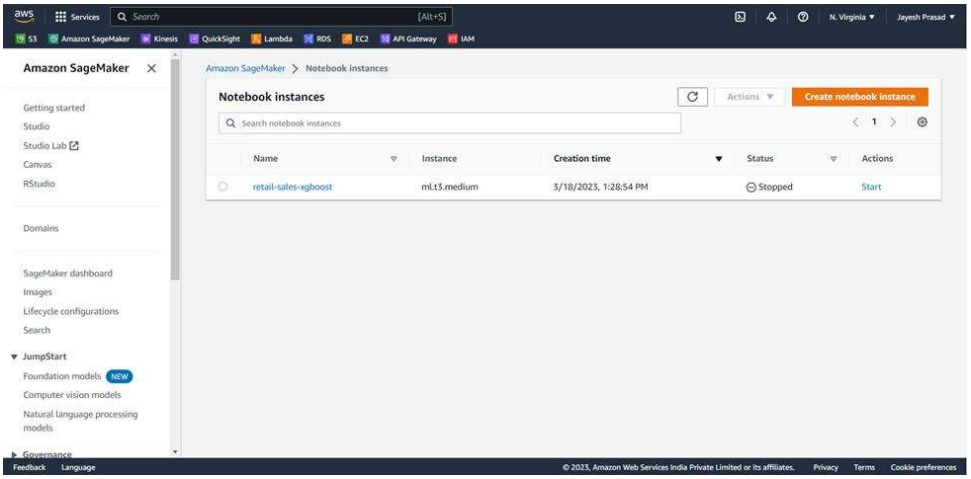




2. Transferred the [Retail Sales Dataset from Kaggle] to AWS S3.

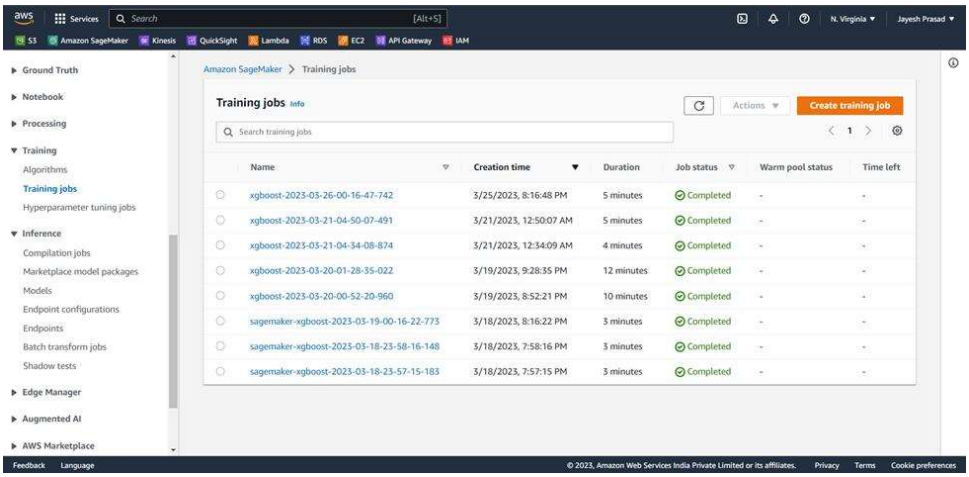


3. Utilized AWS Sagemaker Notebook instances to load the data.



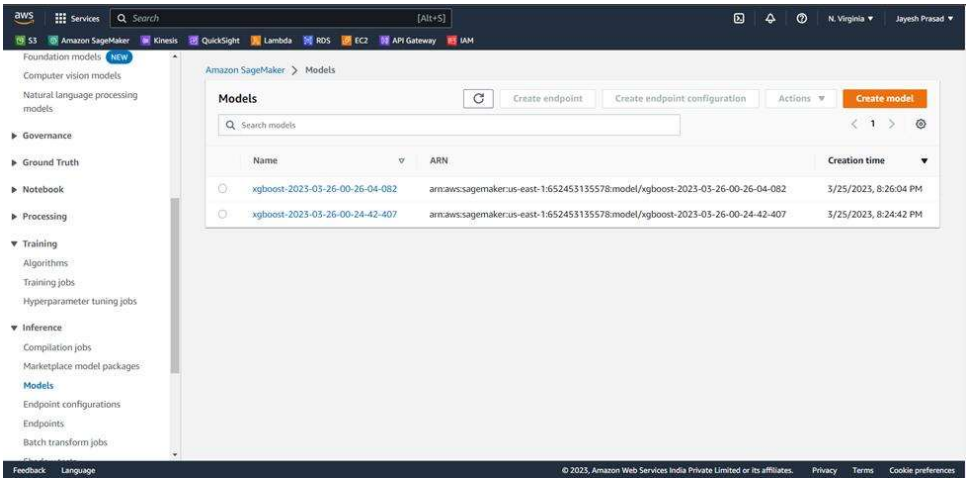
4. Performed Exploratory Data Analysis and preprocessing on the data.

5. Implemented XGBoost prediction model and obtained 90.5% prediction accuracy on the validation data.



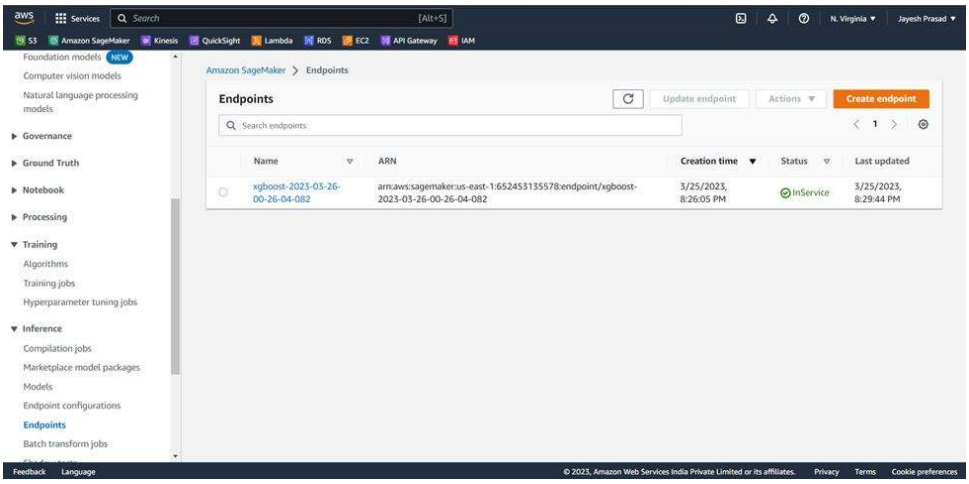
6. Performed Hyperparameter Tuning job to find the best parameters for the given data.

7. Trained the AWS Sagemaker XGBoost model with the best hyperparameters.

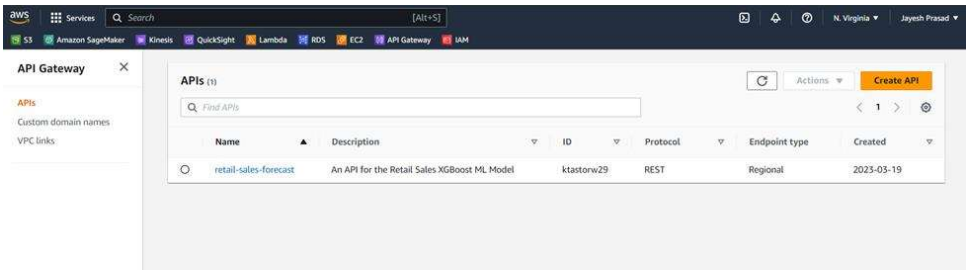


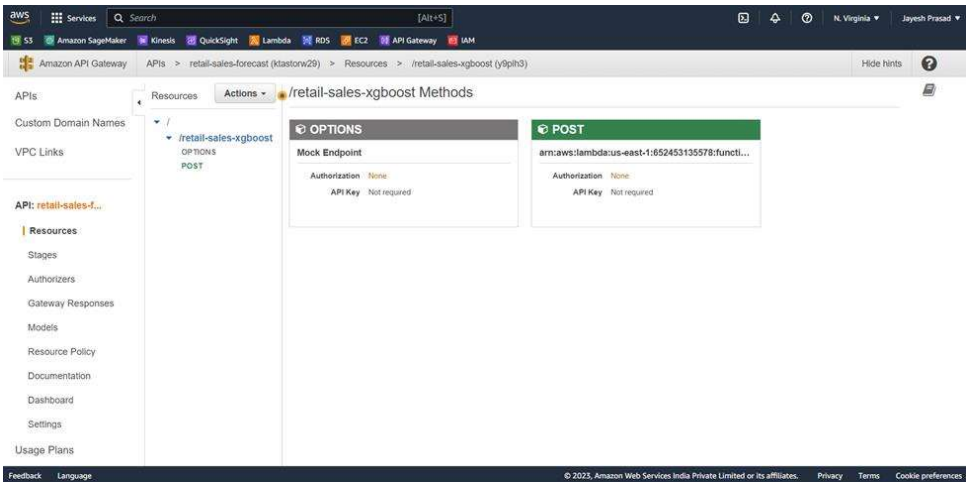
8. Validated the data with test data and received 98.43% Accuracy.

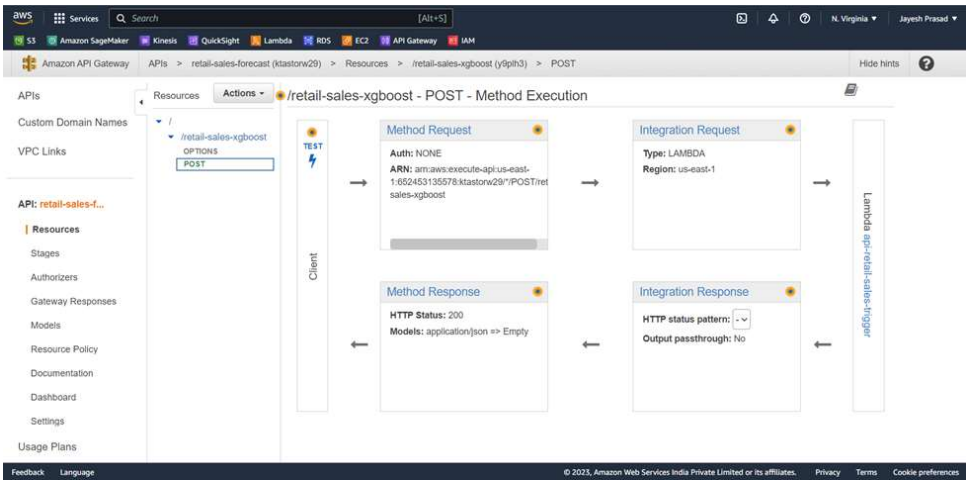
9. Created an Endpoint for the model with a ml.t2.medium instance to be accessed through AWS Lambda.



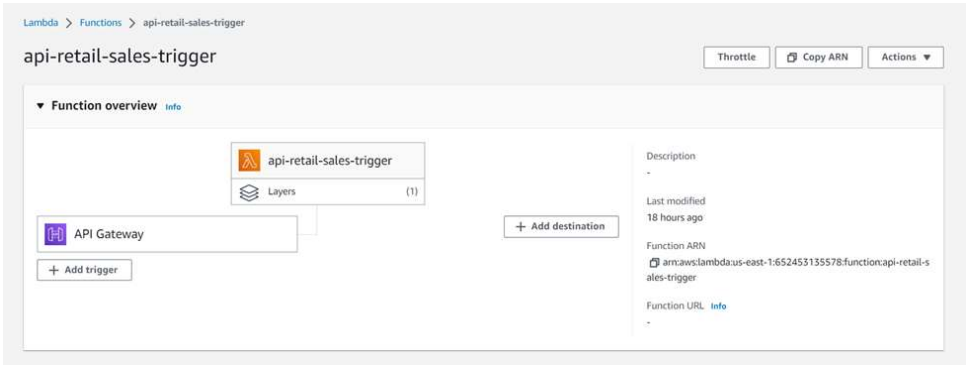
10. Created an API using AWS API Gateway and added a post function to receive input data.

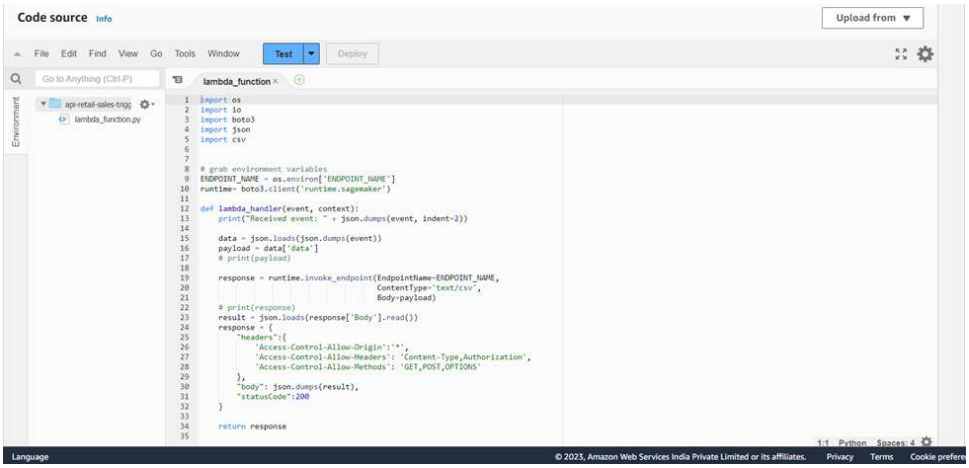


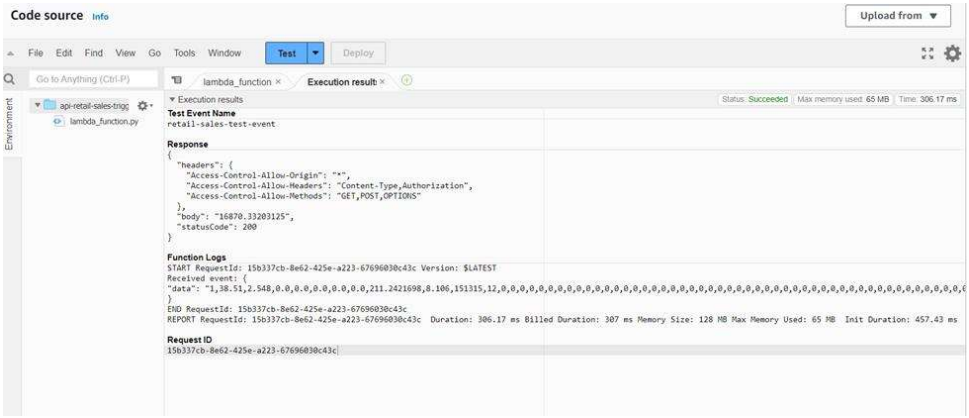




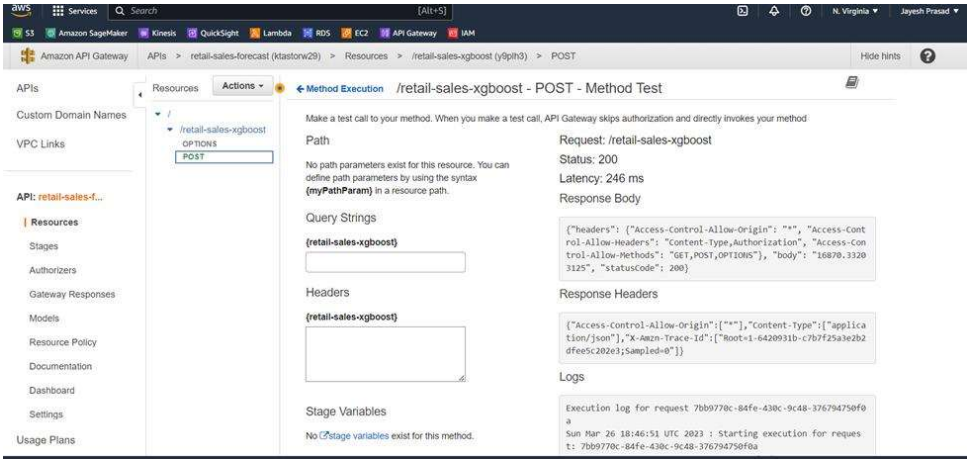
11. Connected AWS Lambda and AWS API Gateway. The API will trigger the model Endpoint with the given input.



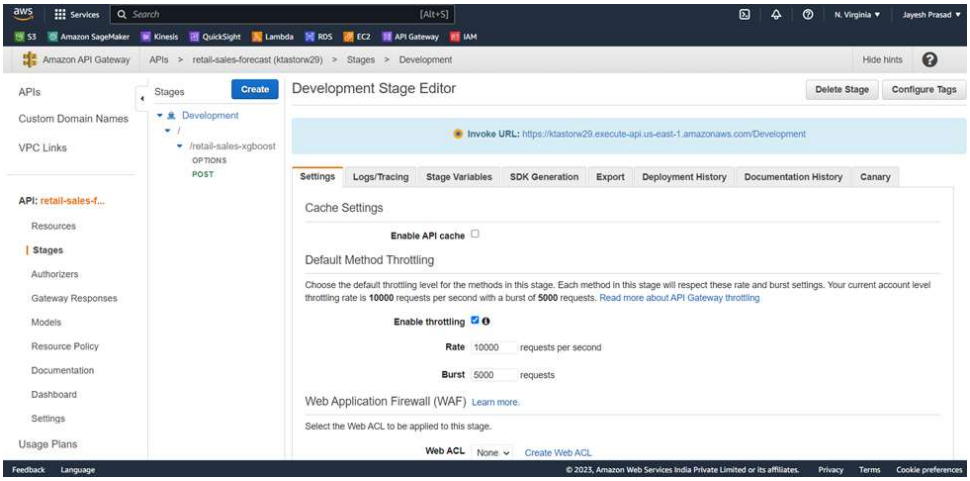




12. Tested the API with a sample input.



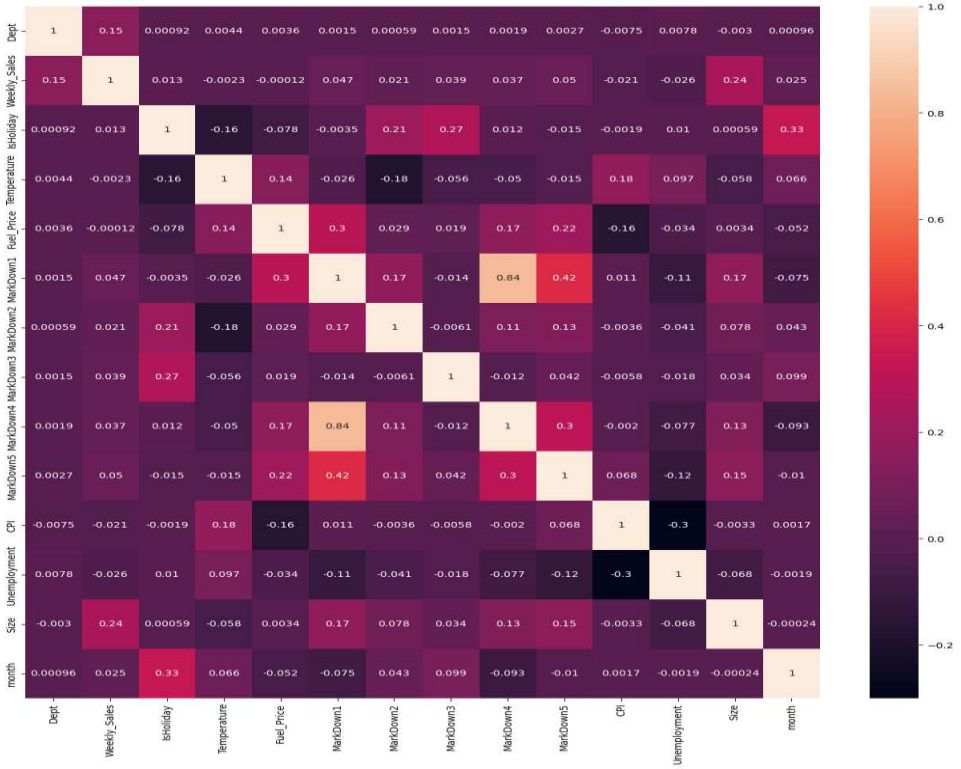
13. Deployed the API to be available for anyone to access the model.



14. Currently Planning the UI design for optimal user engagement to use the API.

Observations

- Identified the patterns and correlations between the features by performing data preprocessing and EDA.



  - The Weekly Sales are positively correlated with Department and Size of the Store.
  - The Markdown parameter values tends to increase with isHoliday parameter and the Month.
- Obtained an impressive prediction accuracy of 90.5% on the validation data by implementing the XGBoost prediction model.
- We trained the AWS Sagemaker XGBoost model with the best hyperparameters and validated with test data, which resulted in an accuracy of 98.43%.
- We have uploaded dataset, notebook and documents in the following GitHub repository:

  https://github.iu.edu/janandan/ecc-project-retail-sales-aws

- Verified the functionality of API with a sample input, after connecting AWS Lambda and AWS API Gateway. Following successful testing, the API is made publicly available for anyone to access the model with the following link:

  https://ktastorw29.execute-api.us-east-1.amazonaws.com/Development/retail-sales-xgboost