

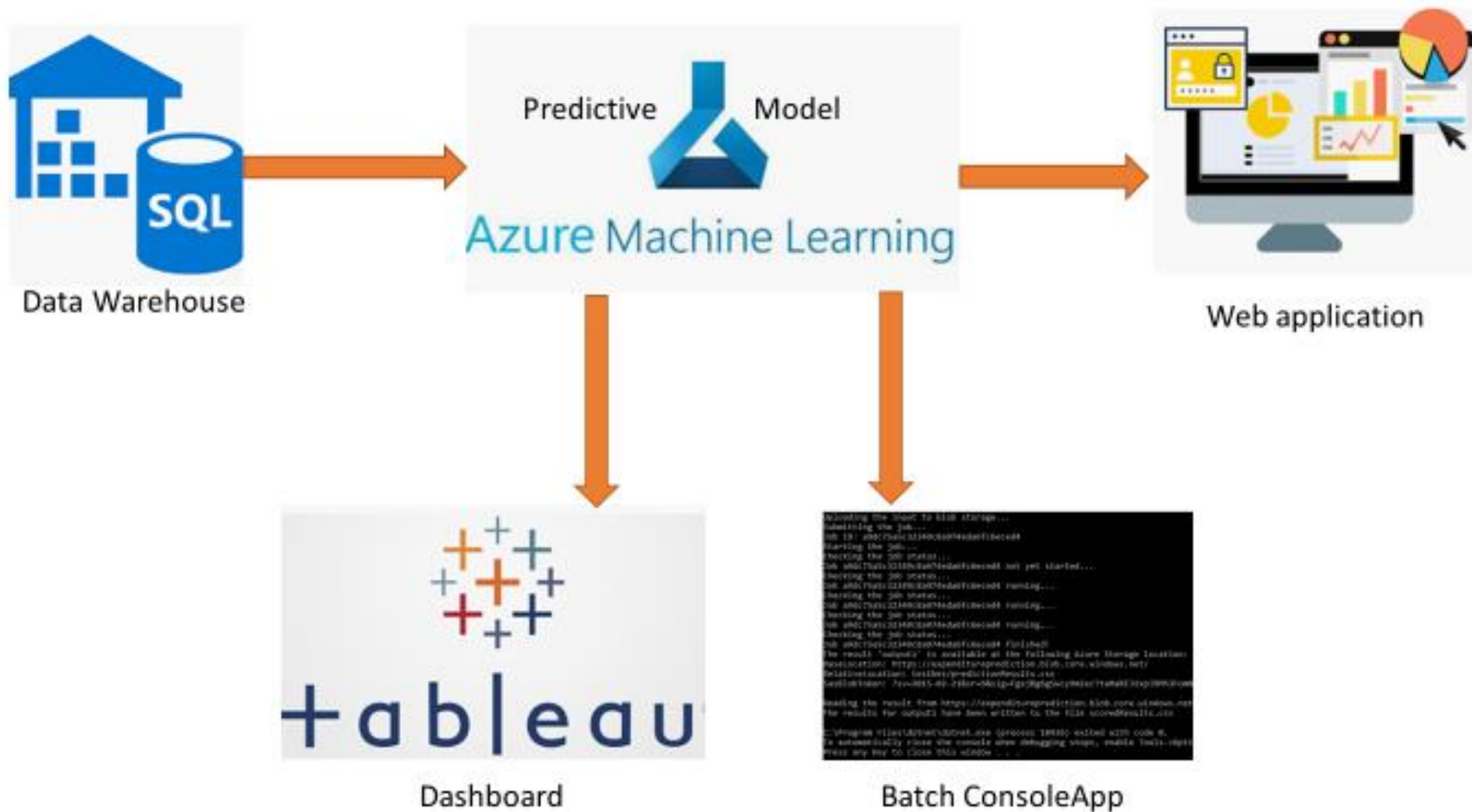


Name: John Onderi Mekubo  
Student ID: 001037854

Programme  
MSc. Big Data and Business Intelligence

Project Title  
Using Predictive Analytics in the Cloud  
to Predict Household Commodity Expenditure

## Prediction Solution Design



# Project Structure



Data Warehouse-Data Storage in Microsoft SQL Server



Prediction Model- Azure Machine Learning Cloud



Prediction Web Application-Using ASP. Net C#-Single Commodity Prediction



Prediction Console Application-To perform batch processing.



Visualization dashboards- developed in tableau and incorporated in web application.

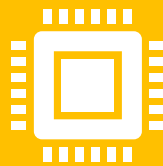
# Data Warehouse



Star Schema model was used



SQL Server Management Studio  
used to create data warehouse  
and its tables.



SQL Server Integration Services  
was used to load data into  
dimension and Fact tables.



# Dataset

---

- The dataset is about UK weekly household commodity expenditure trends for the years, 2015 to 2018. It contains a total of 150 commodity items and 3600 records distributed across 44 categories and 5 different age groups.

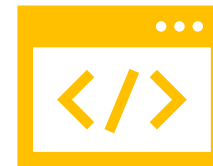
# Selecting Prediction Model



Prediction Model was developed using Azure Machine Learning Cloud platform.



Various Machine learning algorithms were tested i.e. Neural Network regression, Boosted Decision Tree Regression, Decision Forest Regression and Linear Regression.



Linear Regression stood as the ideal algorithm to use. The other algorithms produced results that could not be accepted

# Why Linear Regression Prediction Model



While the other algorithms had higher accuracy score, they were not appropriate in prediction of this dataset as seen in tested results in the next slides. Linear Regression was accurate and was thus chosen as the model to use.

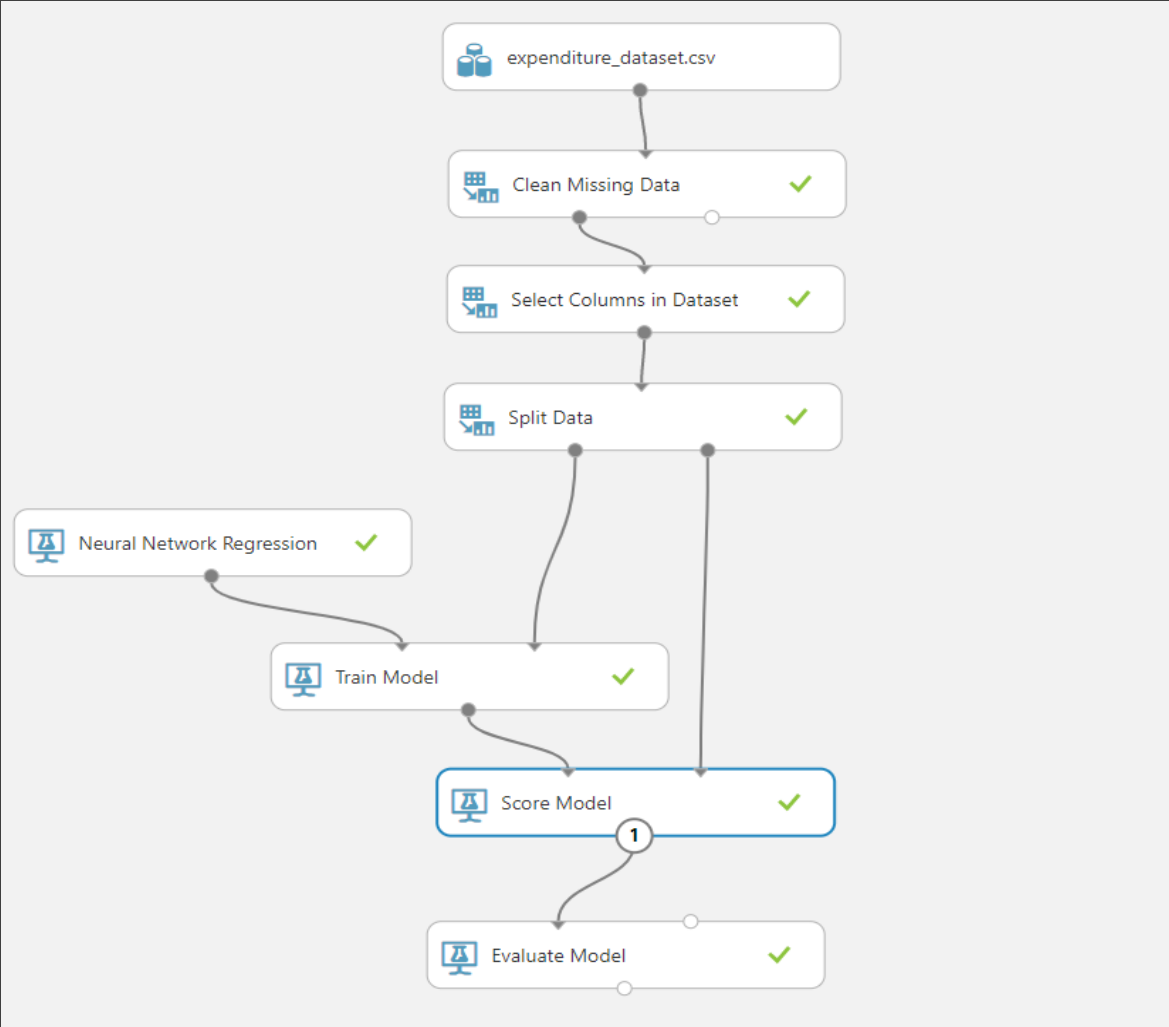


Linear Regression is the ideal algorithm in predicting numerical data.



During Testing of the three prediction models, the other algorithms produced results that could not be accepted

# Testing of Neural Network Regression Model



neural network regression-model -deploy

DASHBOARD CONFIGURATION

General New Web Services Experience [preview](#)

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

ZUNOpGiavnhfwBAdhSMGbE2uN6JoGiVpnMp9/Iy0k3+hUKazTMqdCX6LD8/hOGXMA5CNETkTrana8ZK

Default Endpoint

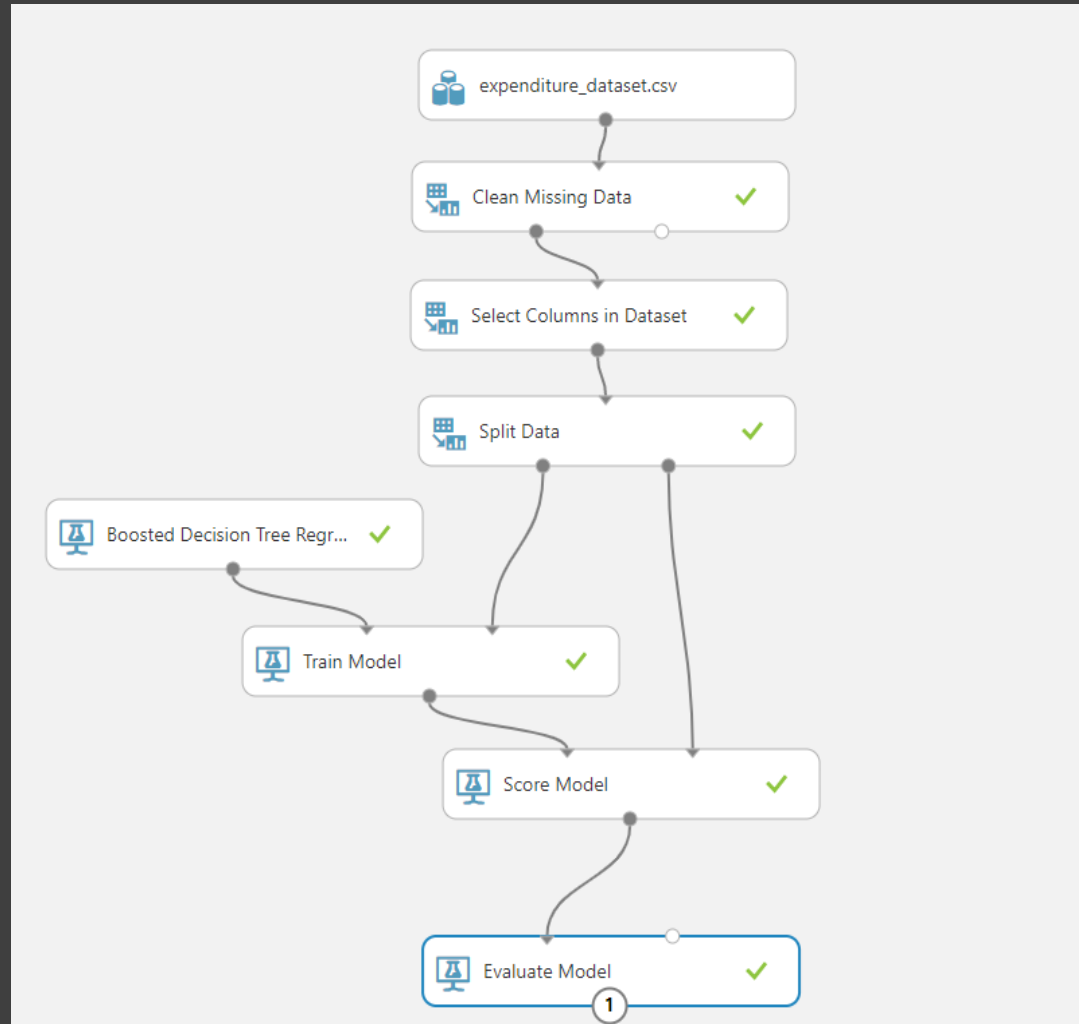
API HELP PAGE	TEST
REQUEST/RESPONSE	<a href="#">Test</a> <a href="#">Test preview</a>
BATCH EXECUTION	<a href="#">Test preview</a>

3 OPERATIONS HAVE COMPLETED

- ✓ 'Neural Network Regression-Model -Deploy' test returned ["Newspapers", "< 30", "2019", "1.87754535675049"]...
- ✓ 'Neural Network Regression-Model -Deploy' test returned ["Newspapers", "< 30", "2022", "23.7462463378906"]...
- ✓ 'Neural Network Regression-Model -Deploy' test returned ["Newspapers", "< 30", "2030", "52.9399604797363"]...



# Testing of Boosted Decision Tree Model



boosted decision tree regression-model -deploy

DASHBOARD CONFIGURATION

General [New Web Services Experience](#) preview

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

vsI2mDV7MDPpBjc1GpEBzz8T3suNFAGZlOCi3myQlPuSK6zZPGoHevRJvxP1yskFP8oeOLiGT7DCWI7/3cj+bQ==

Default Endpoint

API HELP PAGE TEST

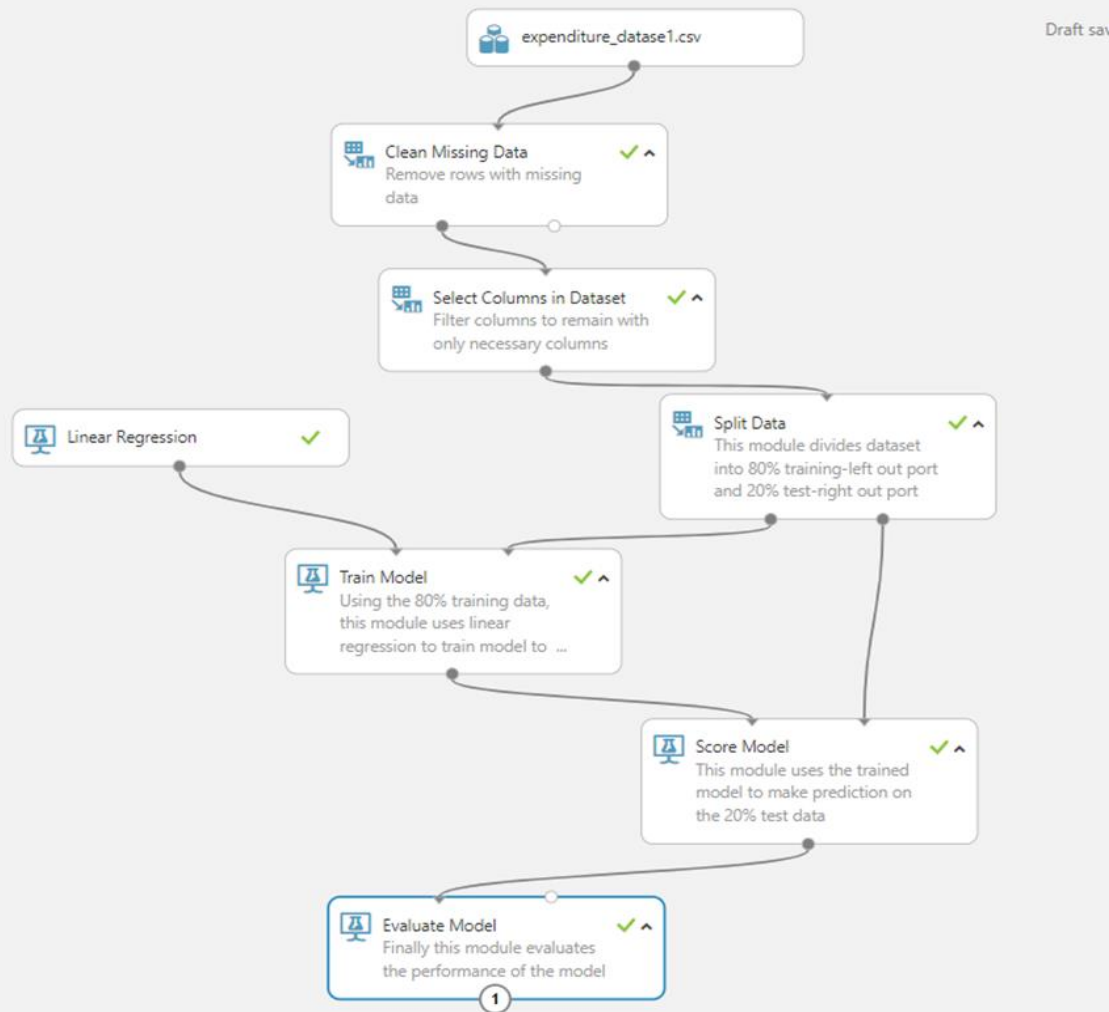
REQUEST/RESPONSE [Test](#) [Test](#) preview

BATCH EXECUTION [Test](#) preview

3 OPERATIONS HAVE COMPLETED

- ✓ 'Boosted Decision Tree Regression-Model -Deploy' test returned ["Newspapers", "< 30", "2019", "1.00615131855011"]...
- ✓ 'Boosted Decision Tree Regression-Model -Deploy' test returned ["Newspapers", "< 30", "2022", "1.00615131855011"]...
- ✓ 'Boosted Decision Tree Regression-Model -Deploy' test returned ["Newspapers", "< 30", "2030", "1.00615131855011"]...

# Testing Linear Regression Model



## linear regression prediction model -deploy

DASHBOARD CONFIGURATION

General [New Web Services Experience](#) preview

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

AMQ6b1YbZDVOKSsK+dwoqrCnXD/GFDkRzX67wussADwCBx5MHhskGC7Md0Is118ofysTyfyi4Me4OTDtswFjlw==

Default Endpoint

[API HELP PAGE](#)

TEST

[REQUEST/RESPONSE](#)

[Test](#)

[Test](#) preview

[BATCH EXECUTION](#)

[Test](#) preview

3 OPERATIONS HAVE COMPLETED

✓ 'Linear Regression Prediction Model -Deploy' test returned ["Newspapers", "< 30", "2019", "1.65180868360662"]...

✓ 'Linear Regression Prediction Model -Deploy' test returned ["Newspapers", "< 30", "2022", "1.81851803114107"]...

✓ 'Linear Regression Prediction Model -Deploy' test returned ["Newspapers", "< 30", "2030", "2.26307629123295"]...

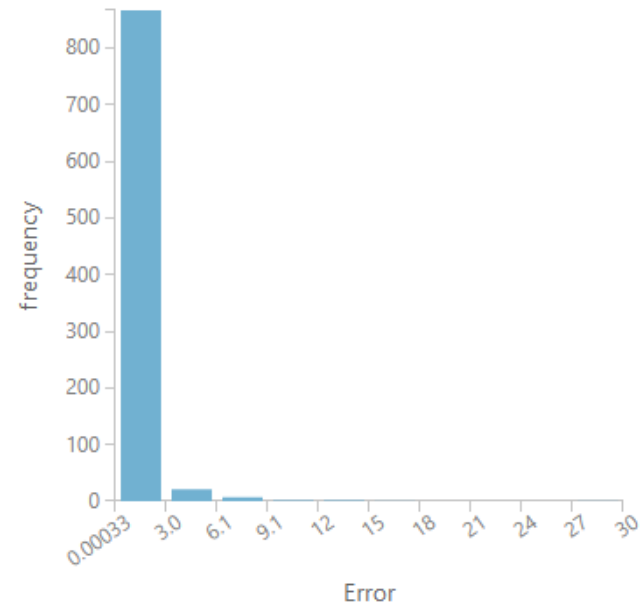
# Neural Network Regression Prediction Model Evaluation Results

Neural Network Regression-Mode > Evaluate Model > Evaluation results

## Metrics

Mean Absolute Error	0.830939
Root Mean Squared Error	1.86819
Relative Absolute Error	0.180349
Relative Squared Error	0.036992
Coefficient of Determination	0.963008

## Error Histogram



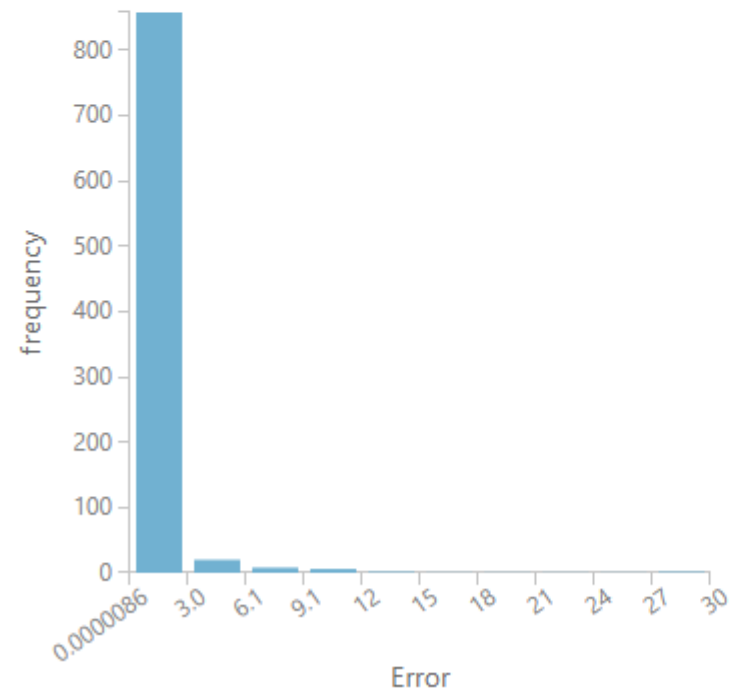
# Boosted Decision Tree Model Evaluation Results

Boosted Decision Tree Regression-Model > Evaluate Model > Evaluation results

## Metrics

Mean Absolute Error	1.073422
Root Mean Squared Error	2.597902
Relative Absolute Error	0.232978
Relative Squared Error	0.071535
Coefficient of Determination	0.928465

## Error Histogram



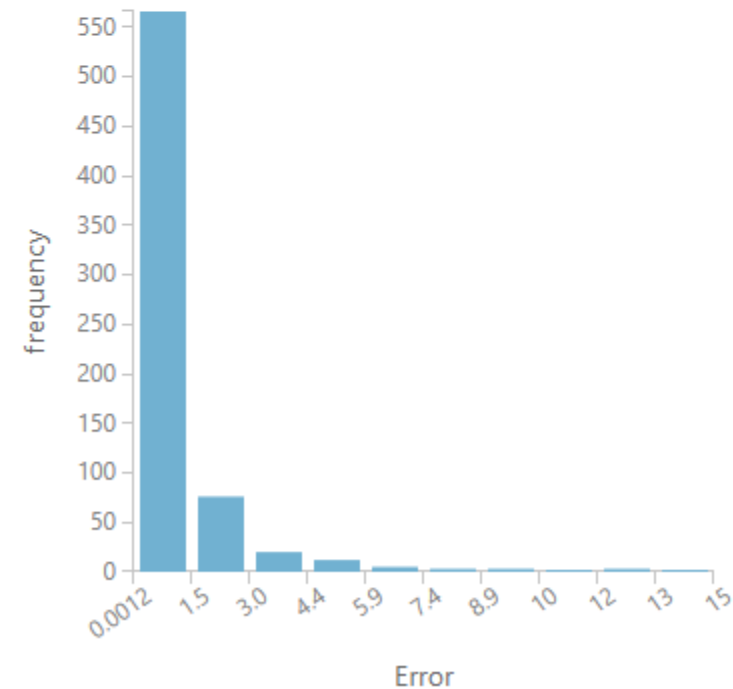
# Linear Regression Model Evaluation Results

Linear Regression Prediction Model > Evaluate Model > Evaluation results

## Metrics

Mean Absolute Error	1.097585
Root Mean Squared Error	2.083877
Relative Absolute Error	0.332845
Relative Squared Error	0.171327
Coefficient of Determination	0.828673

## Error Histogram



localhost/prediction/ ☆ ☆ ✎ ↻ ...

board1 Prediction Application

### Single Commodity Prediction

Commodity

Age Group

Enter Prediction Year

Scored Results

Electricity

30 to 49

2025

12.1160183208733

- The web application allows a user to make prediction of a single commodity by selecting commodity, age group and year of prediction. The scored results are displayed after clicking on prediction button

# Prediction Web Application

# Prediction Console Application

The Console application copies the input csv file to Azure Blob storage and uses the prediction model in Azure Machine Learning to make prediction. The results are stored in Azure Blob storage and can be downloaded as csv file.



```
Starting the job...  
Checking the job status...  
Job a9dc75a5c32349c8a974eda6fc6eced4 not yet started...  
Checking the job status...  
Job a9dc75a5c32349c8a974eda6fc6eced4 running...  
Checking the job status...  
Job a9dc75a5c32349c8a974eda6fc6eced4 running...  
Checking the job status...  
Job a9dc75a5c32349c8a974eda6fc6eced4 running...  
Checking the job status...  
Job a9dc75a5c32349c8a974eda6fc6eced4 finished!  
The result 'output1' is available at the following Azure Storage location:  
BaseLocation: https://expenditureprediction.blob.core.windows.net/  
RelativeLocation: testbes/predictionResults.csv
```

# Tableau Dashboard



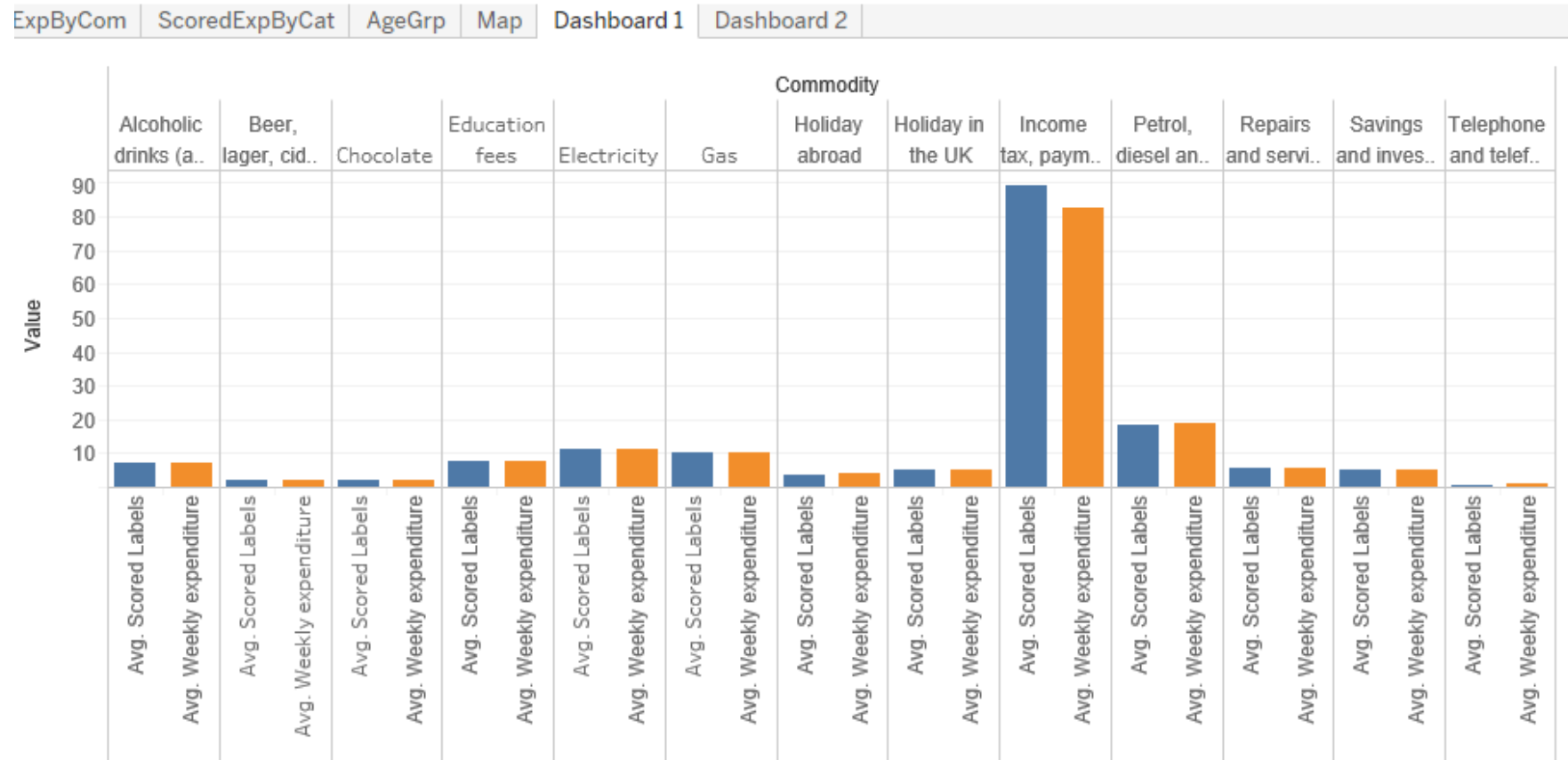
The prediction results file may be used for analytics and visualization to gain more insight into the data



Tableau was used to develop visualization and dashboards that were integrated into the web application



# Dashboard





Thank You