

Apply a Predictive Model in Oracle Analytics

Before you Begin

This lab shows you how to use a CART predictive model, inspect the results of running the data flow to create the CART model, apply a predictive model to your dataset, and visualize the impact in a workbook.

Background

You can create a different predictive model using a classification and regression tree (CART) for numeric prediction with the random sample donation dataset. The CART predictive model classifies all of the records in the dataset based on inputs, and assigns an expected record for each class. In this lab, you use the CART model to predict the total donations.

You also score a dataset using the trained model by applying the predictive model. You run analysis and create visualizations using the predictive measure along with other metrics, dimensions, or dataset mash-ups.

This is the third lab in *Train and Apply Predictive Models* in Oracle Analytics. Read the labs in the order listed.

- Create a Random Sample Dataset and Train a Prediction Model
- Inspect and Modify the Prediction Model
- Apply a Predictive Model


What Do You Need?


- Oracle Analytics

When using Oracle Analytics Desktop, you must install machine learning (DVML) to use Diagnostics Analytics (Explain), Machine Learning Studio, or advanced analytics.



- Access to the following:
 - **elastic_train_df** data flow
 - **sample_donation_data** dataset
 - **donations_random_sample** workbook

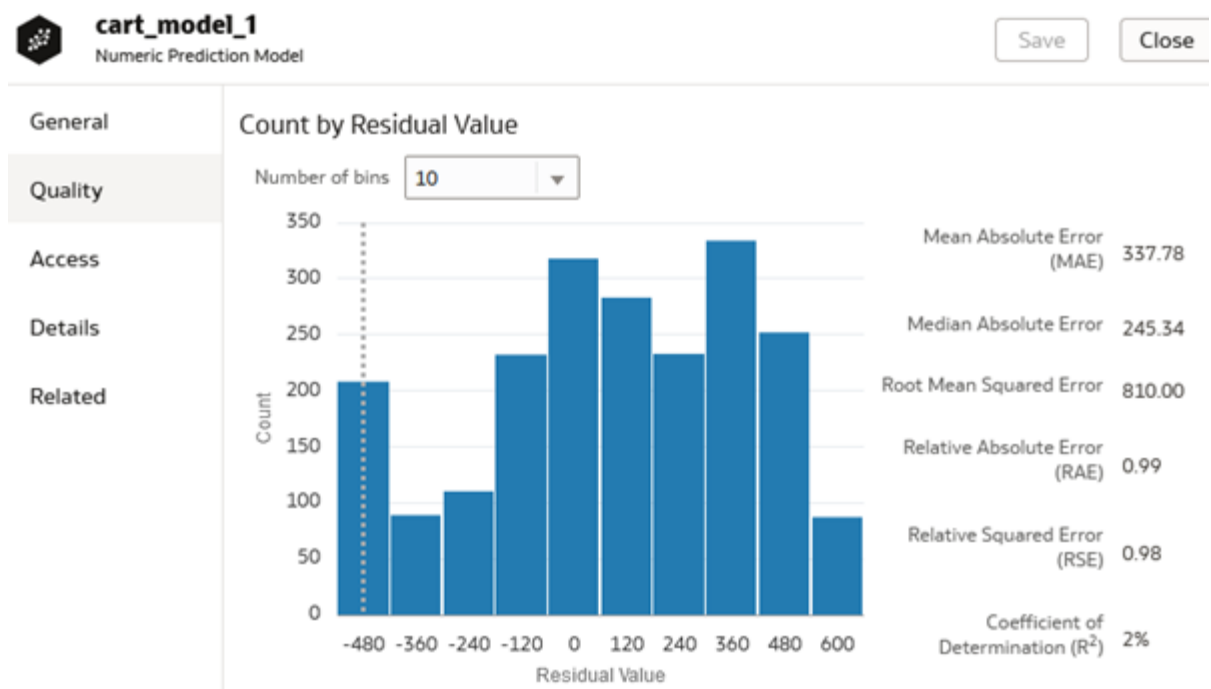
Modify the Predictive Model

1. On the Home page, click **Data**, enter **elastic_train_df** in the Search bar, and then click **Search**.
2. In the **elastic_train_df**, click the **Actions menu** , and then select **Open**.
3. In the data flow, click the **Train Numeric Prediction** step.
4. In Train Numeric Prediction, click **Elastic Net Linear Regression for model training** in the Model Training Script row.

5. In Select Train Numeric Prediction Model Script, click **CART for Numeric Prediction training**, and then click **OK**.
6. In Train Numeric Prediction, click **Select a column** in the Target row, and then select **TOTAL_DONATIONS** from Available Data.
7. Click the **Save Model** step. In Save Model, enter cart_model_1. Click **Save**.
8. Click **Run Data Flow** .


Review the CART Model Related Datasets

1. Click **Go back** . On the Home page, click **Machine Learning**.
2. In the cart_model_1, click the **Actions menu** , and then select **Inspect**.
3. In cart_model_1, click **Quality** to see the distribution of residual values and overall statistics.



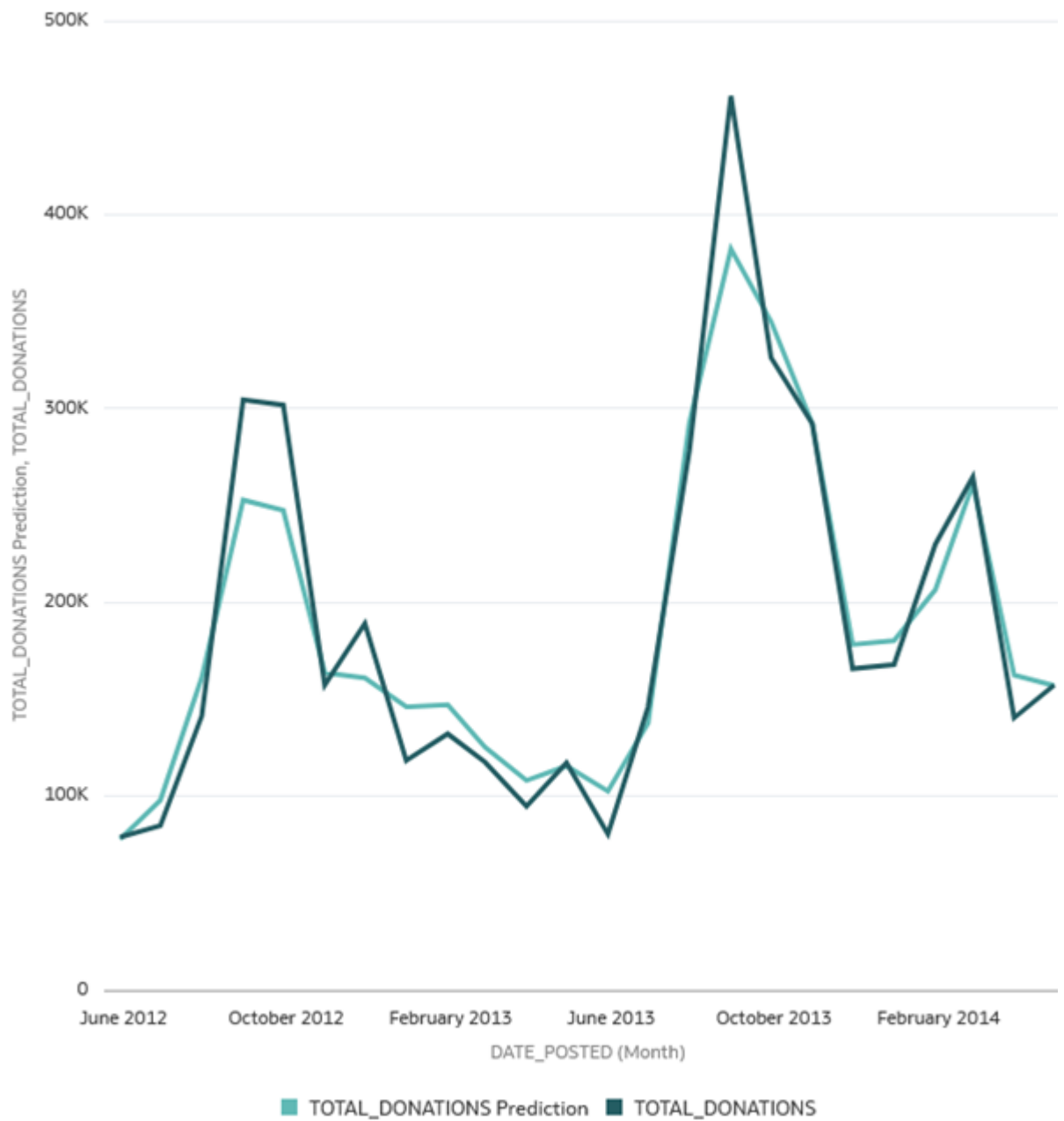
4. Click **Related** to view the generated datasets.
5. Click **Close**.


Add the CART Scenario to the Workbook

1. On the Home page, enter donations_random_sample workbook, and then click **Search**.
2. In the donations_random_sample workbook, click the **Actions menu** , and then select **Open**.
3. In the Data panel, right-click **elastic_model_1**, and then select **Remove From Workbook**. In Remove Scenario, click **Yes**.

The TOTAL_DONATIONS by POSTED_DATE line visualization refreshes to remove the prediction line related to the scenario.

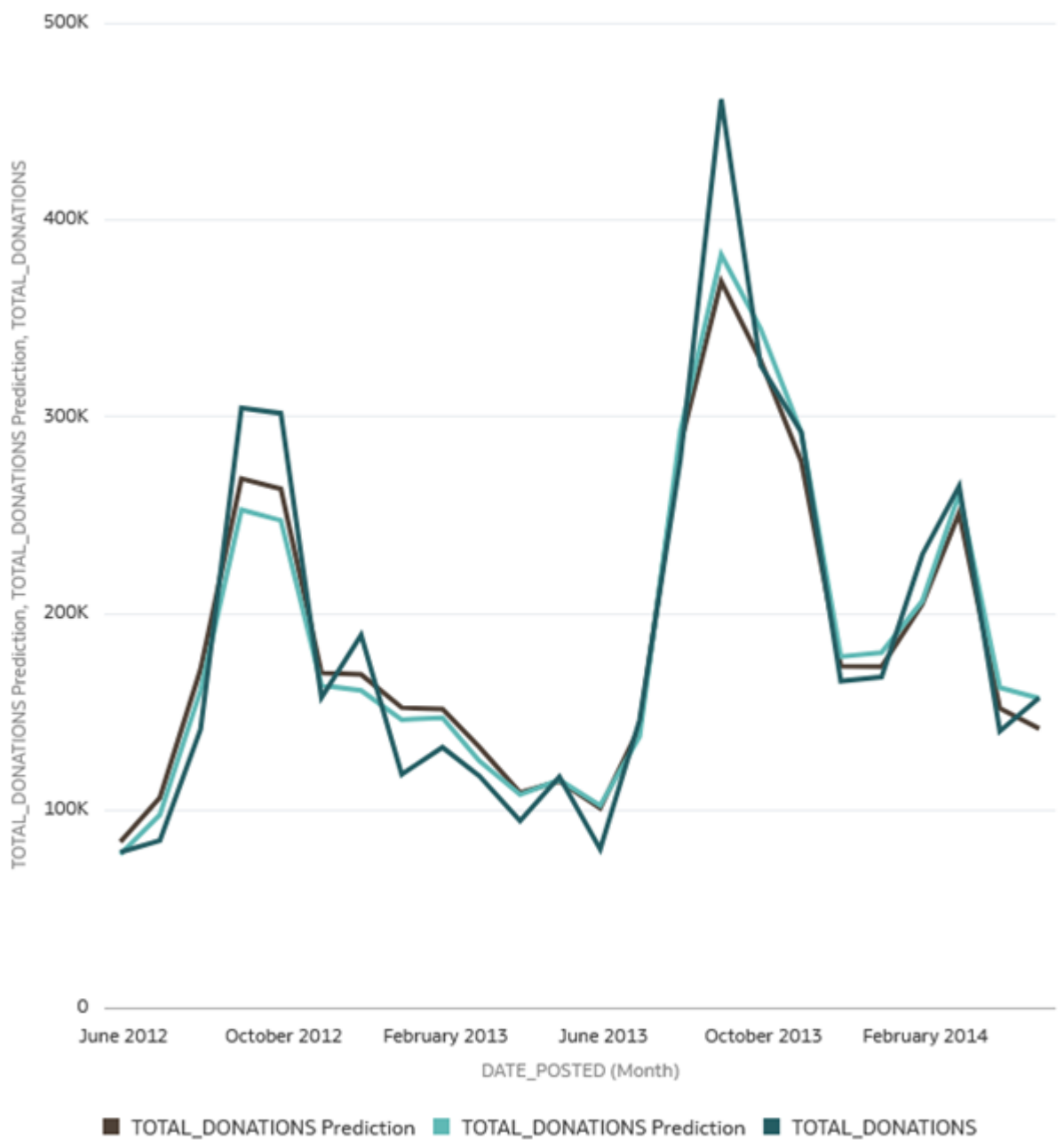
TOTAL_DONATIONS Prediction, TOTAL_DONATIONS by DATE_POSTED (Month)



4. In the Data panel, click **Add** , and then select **Create Scenario**.
5. In Create Scenario, select **cart_model_1**, and then click **OK**.
6. In the Data panel, expand the **cart_model_1** node. Select **TOTAL_DONATIONS**, and then drag it to **Values (Y-Axis)** in the Grammar panel.


The CART numeric prediction model produces results similar to the elastic model.

TOTAL_DONATIONS Prediction, TOTAL_DONATIONS Prediction, TOTAL_DONATIONS by DATE_P...



7. Click **Save**.

Review CART Numeric Model Segments

1. Click **Add Canvas** . In the Data panel under the cart_model_1 node, select and drag **Segments** to **Canvas 2**.

The CART model segment objects show the leaf branches of the tree.

Segments


Segments
RESOURCE_TYPE NOT IN ['Trips', 'Technology'];PRIMARY_FOCUS_SUBJECT IN ['Music', 'Foreign Languages', 'Health & Life Science', 'Extracurricular'] AND PRIMARY_FOCUS_SUBJECT NOT IN ['Civics & Government', 'Applied Sciences', 'Early Development', 'Environmental Science']
RESOURCE_TYPE NOT IN ['Trips', 'Technology'];PRIMARY_FOCUS_SUBJECT NOT IN ['Applied Sciences', 'Health & Life Science', 'Extracurricular', 'Foreign Languages', 'Environmental Science', 'Music', 'Civics & Government', 'Early Development']
RESOURCE_TYPE NOT IN ['Trips'];SCH_COUNTY IN ['Harrisonburg City'];PRIMARY_FOCUS_SUBJECT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']
RESOURCE_TYPE NOT IN ['Trips'];SCH_COUNTY NOT IN ['Harrisonburg City'];GRADE_LEVEL IN ['D: Grades 9-12'];PRIMARY_FOCUS_SUBJECT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']
RESOURCE_TYPE NOT IN ['Trips'];SCH_COUNTY NOT IN ['Harrisonburg City'];GRADE_LEVEL NOT IN ['D: Grades 9-12'];PRIMARY_FOCUS_SUBJECT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']
SCH_CITY IN ['Murrieta'];RESOURCE_TYPE IN ['Trips']
SCH_CITY NOT IN ['Murrieta'];RESOURCE_TYPE IN ['Trips'];GRADE_LEVEL IN ['D: Grades 9-12']
SCH_CITY NOT IN ['Murrieta'];RESOURCE_TYPE IN ['Trips'];GRADE_LEVEL NOT IN ['D: Grades 9-12']
SECONDARY_FOCUS_SUBJECT IN ['Social Sciences', 'Applied Sciences'];RESOURCE_TYPE IN ['Technology'] AND RESOURCE_TYPE NOT IN ['Trips'];PRIMARY_FOCUS_SUBJECT NOT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']
SECONDARY_FOCUS_SUBJECT NOT IN ['Social Sciences', 'Applied Sciences'];RESOURCE_TYPE IN ['Technology'] AND RESOURCE_TYPE NOT IN ['Trips'];PRIMARY_FOCUS_SUBJECT NOT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']

2. In the Data panel under the cart_model_1 node, select and drag **TOTAL_DONATIONS** to **Rows** in the Grammar panel.

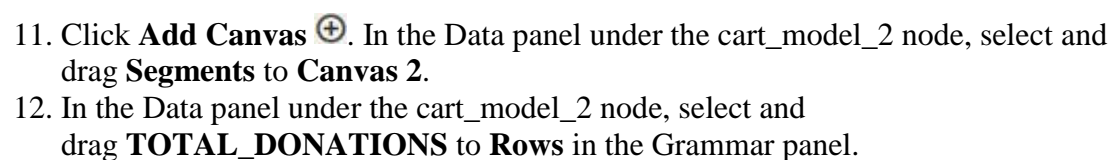
The table shows the distribution of donations to the CART branches.

Segments	TOTAL_DONATIONS Prediction
RESOURCE_TYPE NOT IN ['Trips', 'Technology'];PRIMARY_FOCUS_SUBJECT IN ['Music', 'Foreign Languages', 'Health & Life Science', 'Extracurricular'] AND PRIMARY_FOCUS_SUBJECT NOT IN ['Civics & Government', 'Applied Sciences', 'Early Development', 'Environmental Science']	255,709.71
RESOURCE_TYPE NOT IN ['Trips', 'Technology'];PRIMARY_FOCUS_SUBJECT NOT IN ['Applied Sciences', 'Health & Life Science', 'Extracurricular', 'Foreign Languages', 'Environmental Science', 'Music', 'Civics & Government', 'Early Development']	1,811,532.89
RESOURCE_TYPE NOT IN ['Trips'];SCH_COUNTY IN ['Harrisonburg City'];PRIMARY_FOCUS_SUBJECT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']	7,982.76
RESOURCE_TYPE NOT IN ['Trips'];SCH_COUNTY NOT IN ['Harrisonburg City'];GRADE_LEVEL IN ['D: Grades 9-12'];PRIMARY_FOCUS_SUBJECT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']	183,537.12
RESOURCE_TYPE NOT IN ['Trips'];SCH_COUNTY NOT IN ['Harrisonburg City'];GRADE_LEVEL NOT IN ['D: Grades 9-12'];PRIMARY_FOCUS_SUBJECT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']	538,256.04
SCH_CITY IN ['Murrieta'];RESOURCE_TYPE IN ['Trips']	10,280.83
SCH_CITY NOT IN ['Murrieta'];RESOURCE_TYPE IN ['Trips'];GRADE_LEVEL IN ['D: Grades 9-12']	58,618.53
SCH_CITY NOT IN ['Murrieta'];RESOURCE_TYPE IN ['Trips'];GRADE_LEVEL NOT IN ['D: Grades 9-12']	34,833.77
SECONDARY_FOCUS_SUBJECT IN ['Social Sciences', 'Applied Sciences'];RESOURCE_TYPE IN ['Technology'] AND RESOURCE_TYPE NOT IN ['Trips'];PRIMARY_FOCUS_SUBJECT NOT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences']	124,812.39

Modify the CART Model

1. Click **Go back** . In Save Changes, click **Don't Save**.
2. On the Home page, enter elastic_train_df, and then click **Search**.
3. In the elastic_train_df, click the **Actions** menu, and then select **Open**.
4. In the data flow, click the **Train Numeric Prediction** step.
5. In **Maximum Depth**, click the up arrow to increase the value to **7**.
6. Click the **Save Model** step. In Save Model, enter cart_model_2.
7. Click **Run Data Flow**.

- The results are similar to the other prediction models.





Increasing the maximum depth in cart_model_2 provides more granular leaf levels in the tree.

Segments, TOTAL_DONATIONS Prediction


Segments	TOTAL_DONATIONS Prediction ▼
PRIMARY_FOCUS_AREA NOT IN ['Music & The Arts', 'Math & Science'];RESOURCE_TYPE IN ['Technology'] AND RESOURCE_TYPE NOT IN ['Trips'];SCH_STATEZIP NOT IN ['CA-92870'];PRIMARY_FOCUS_SUBJECT NOT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences'];SECONDARY_FOCUS_SUBJECT NOT IN ['Social Sciences', 'Applied Sciences']	967,671.32
TCHR_PREFIX IN ['Ms.', 'Dr.', 'Mr.'];RESOURCE_TYPE NOT IN ['Technology', 'Trips'];PRIMARY_FOCUS_SUBJECT NOT IN ['Applied Sciences', 'Health & Life Science', 'Music', 'Civics & Government', 'Extracurricular', 'Foreign Languages', 'Early Development', 'Environmental Science'];DATE_POSTED NOT IN ['2012-09-25 00:00:00.000', '2013-09-03 00:00:00.000']	966,468.79
TCHR_PREFIX NOT IN ['Ms.', 'Dr.', 'Mr.'];RESOURCE_TYPE NOT IN ['Technology', 'Trips'];PRIMARY_FOCUS_SUBJECT NOT IN ['Applied Sciences', 'Health & Life Science', 'Music', 'Civics & Government', 'Extracurricular', 'Foreign Languages', 'Early Development', 'Environmental Science'];SCH_COUNTY NOT IN ['Davis', 'Vermillion', 'Fayette', 'Martin', 'Rockingham', 'Wyandotte', 'Henry', 'Bartow', 'Noble', 'Riverside', 'Philadelphia', 'Chatham', 'St Tammany', 'Broome']	810,428.03
PRIMARY_FOCUS_AREA IN ['Music & The Arts', 'Math & Science'];RESOURCE_TYPE IN ['Technology'] AND RESOURCE_TYPE NOT IN ['Trips'];PRIMARY_FOCUS_SUBJECT NOT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences'];SECONDARY_FOCUS_SUBJECT NOT IN ['Social Sciences', 'Applied Sciences'];SCH_COUNTY NOT IN ['Franklin', 'Chatham', 'Accomack']	500,944.45
RESOURCE_TYPE NOT IN ['Technology', 'Trips', 'Other'];POVERTY_LEVEL IN ['C: High poverty', 'D: Highest poverty'];PRIMARY_FOCUS_SUBJECT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences'];SCHOOL_ID NOT IN ['e8599595318aef33252c14a7adb0476d', '88f1243408b252d5f873a08e6a27e210'];GRADE_LEVEL NOT IN ['D: Grades 9-12']	167,673.24
RESOURCE_TYPE IN ['Technology', 'Other'] AND RESOURCE_TYPE NOT IN ['Trips'];PRIMARY_FOCUS_SUBJECT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences'];SECONDARY_FOCUS_SUBJECT NOT IN ['Social Sciences', 'Mathematics'];GRADE_LEVEL NOT IN ['C: Grades 6-8', 'D: Grades 9-12']	130,376.42
RESOURCE_TYPE IN ['Technology'] AND RESOURCE_TYPE NOT IN ['Trips'];STUDENTS_REACHED <= 300.4;PRIMARY_FOCUS_SUBJECT NOT IN ['Early Development', 'Environmental Science', 'Civics & Government', 'Applied Sciences'];SECONDARY_FOCUS_SUBJECT IN ['Social Sciences', 'Applied Sciences'];SCH_COUNTY NOT IN ['San Mateo']	92,441.48

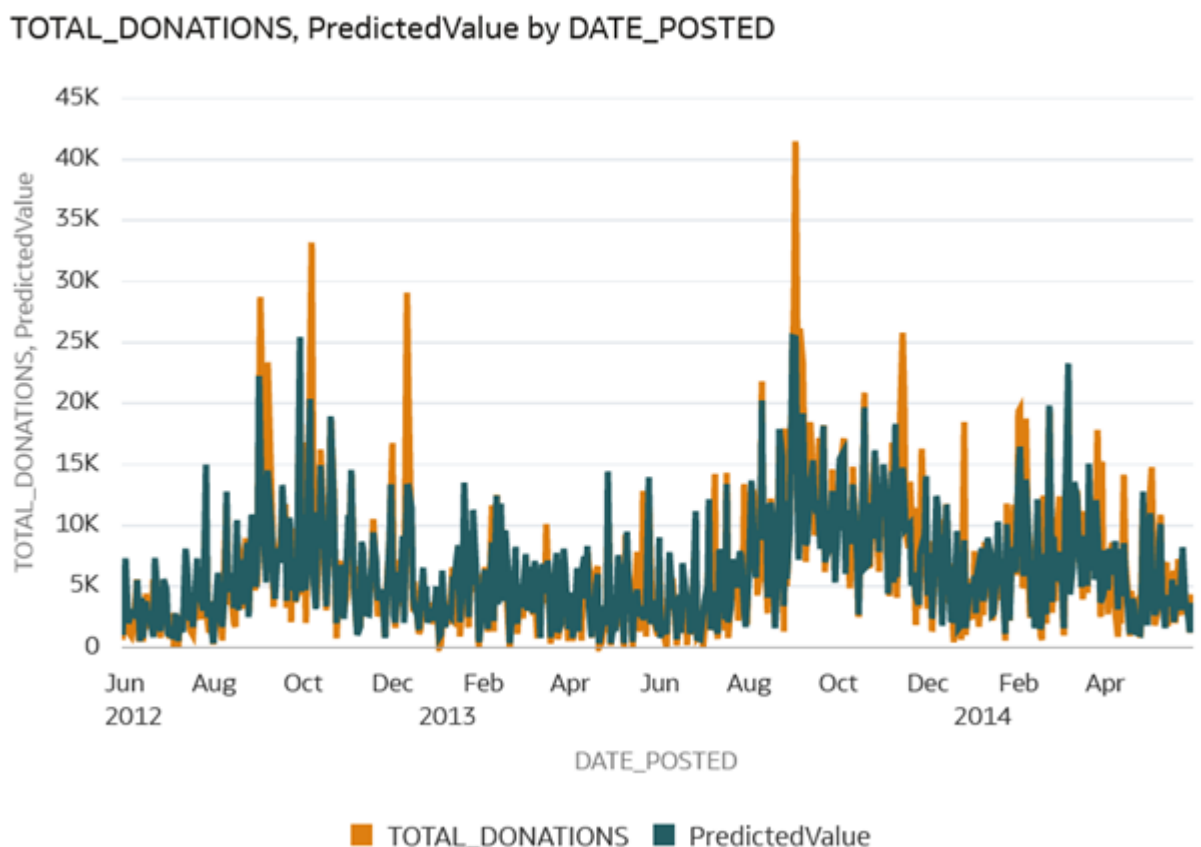
Score a Dataset

1. Click **Go back** . On the Home page, click **Create**, and then select **Data Flow**. In Add Dataset, select **sample_donation_data**, and then click **Add**.
2. Double-click **Apply Model** in the Data Flow Steps panel. In Select Model, click **cart_model_1**, and then click **OK**.

3. Double-click **Save Dataset** to add it to the data flow. In Save Dataset, enter scored_donation_dataset in **Name**.
4. Click **Save**, and select **Save As**. In Save Data Flow As, enter scored_donation_dataset_df, and then click **OK**.
5. Click **Run Data Flow** .

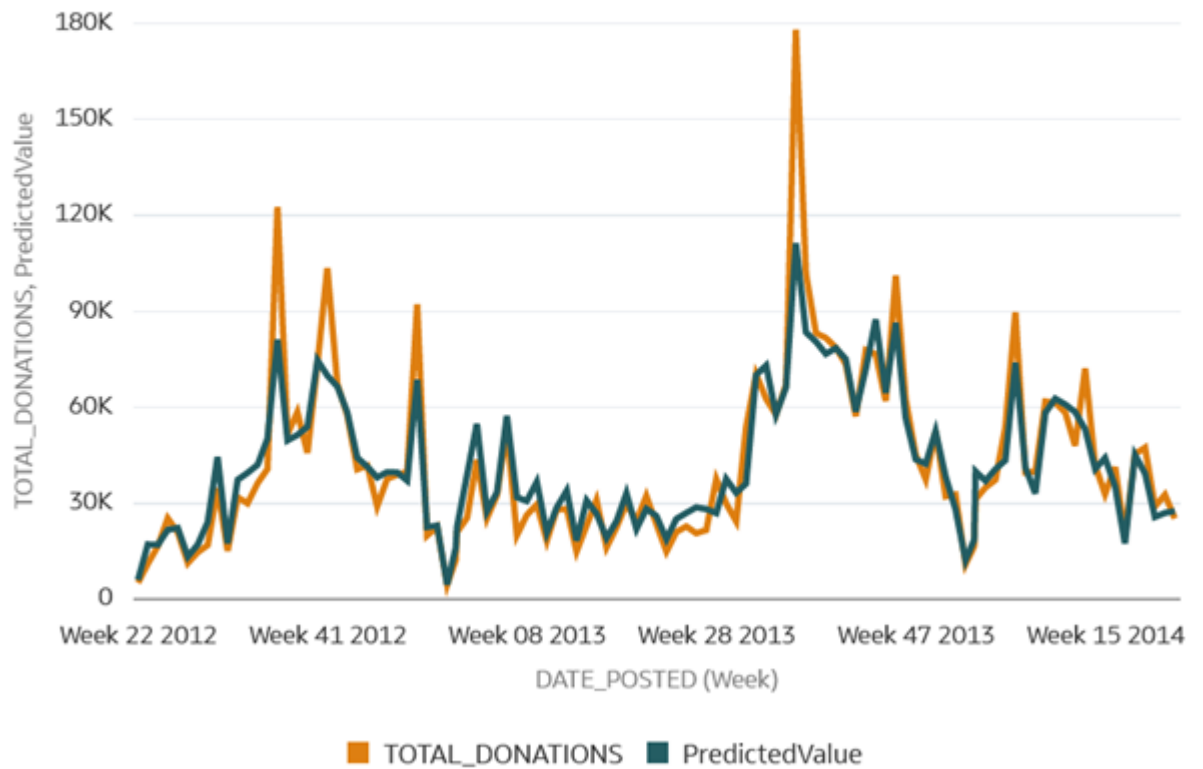
Create a Workbook with the Scored Dataset

1. Click **Go back** . On the Home page, click **Create**, and then click **Workbook**.
2. In Add Dataset, click **scored_donation_dataset**, and then click **Add to Workbook**.
3. In the Data panel, hold down the **Ctrl** key select **PredictedValue** and **TOTAL_DONATIONS** and **DATE_POSTED**. Right-click, select **Pick Visualization**, and then select **Line**.
4. Select **TOTAL_DONATIONS**, and move it to **Values (Y-Axis)** in the Grammar panel.



5. Right-click **DATE_POSTED** in the Grammar panel, select **Show by**, and then select **Week**.

TOTAL_DONATIONS, PredictedValue by DATE_POSTED (Week)



6. Select **SCH_METRO** in the Data panel, and drag it to **Trellis Columns** in the Grammar panel.

TOTAL_DONATIONS, PredictedValue by DATE_POSTED (Week), SCH_METRO

