## Explore Data Flows in Oracle Analytics

Before You Begin

This lab shows you how to create a data flow with two datasets, add columns, transform data in the columns, and create new columns using expressions.

Background

In the tutorial, you create a data flow with a spreadsheet file containing sample school donation data (donation.xslx) and a postal code statistics (zip\_stats.xslx) data file. In the data flow, you modify the donation dataset by filtering data in a column, and then select and add columns. From the two datasets, you create a curated dataset that you can use in analyses.

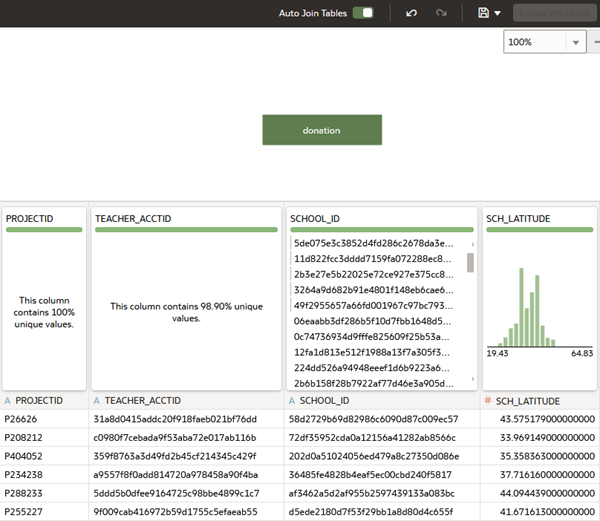
What Do You Need?

* Access to Oracle Analytics Cloud or Oracle Analytics Desktop
* Download the following source files:
  + donation.xlsx
  + zip\_stats.xlsx

Create a Dataset

In this section, you create a dataset using the donation.xlsx file. When numerical data is loaded, it is treated as a measure. You learn how change the Treat as value for numerical columns that are attributes.

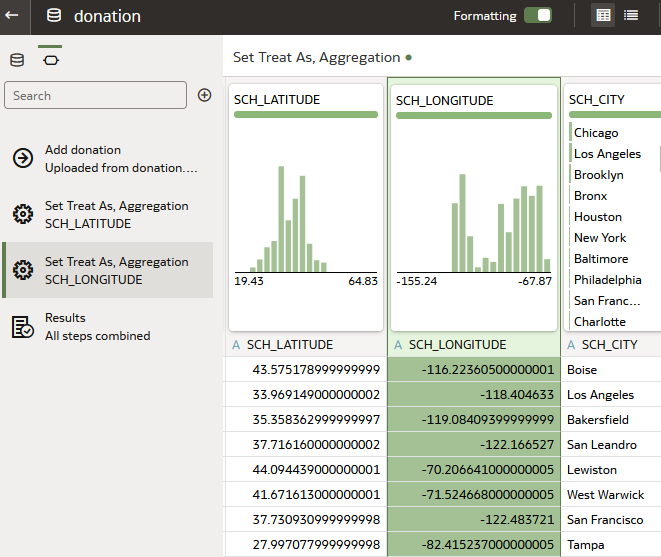
1. Sign in to Oracle Analytics.
2. On the Home page, click **Create**, and then click **Dataset**.
3. In Create Dataset, click **Drop data file here or click to browse**. In File Upload, select the donation.xlsx file, and then click **Open**.
4. In Create Dataset Table from donation.xlsx, click **OK**.



1. Click the donation tab.

Description of donation_tab.png follows

1. Scroll to the **SCH\_LATITUDE** column, click **Measure** measure icon, and then click **Attribute**.
2. Scroll to the **SCH\_LONGITUDE** column, click **Measure** measure icon, and then select **Attribute**.
3. Click **Save** Save icon. In Save Dataset As, enter donation in **Name**, and then click **Save**.

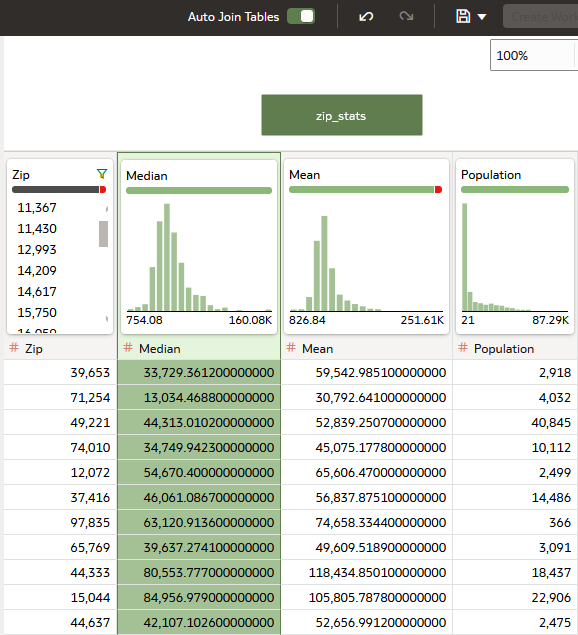


1. In the donation dataset page, click **Go back** Back icon.

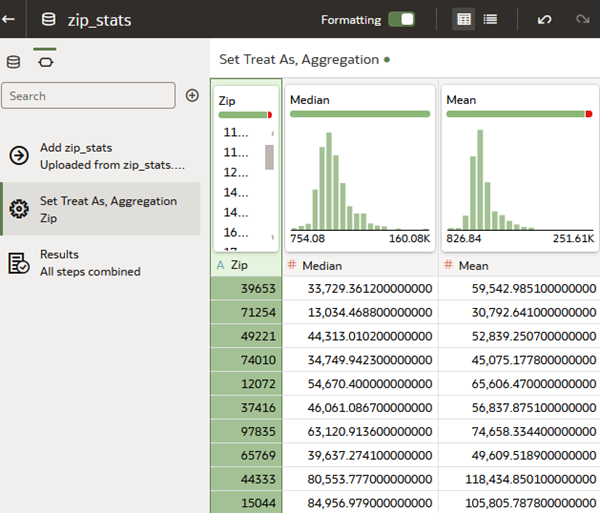
Add a Second Dataset

In this section, you create a dataset that provides demographics to add to the donation dataset.

1. On the Home page, click **Create**, and then click **Dataset**.
2. In Create Dataset, click **Drop data file here or click to browse**. In File Upload, select the zip\_stats.xlsx file, and then click **Open**.
3. In Create Dataset Table from zip\_stats.xlsx, click **OK**.



1. Click the **zip\_stats** tab. In zip\_stats, select the **Zip** column, click **Measure** Measure icon, and then select **Attribute**.
2. In the zip column, click **Options** Options icon, and then click **Convert to Text**.
3. Click **Save** Save icon. In Save Dataset As, enter zip\_stats in **Name**, and then click **Save**.

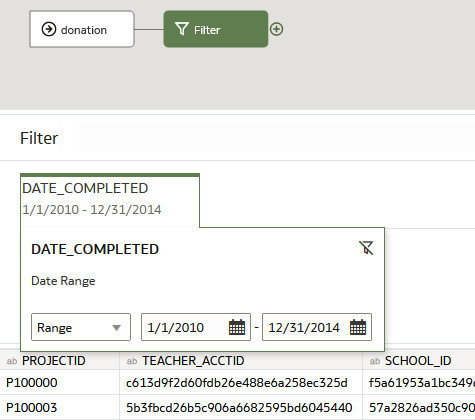


1. Click **Go back** Back icon.

Create a Data Flow

In this section, you create a data flow and implement a filter step to remove the donation records that don't have a value in the DATE\_COMPLETED column.

1. On the Home page, click **Create**, and then click **Data Flow**.
2. In Add Dataset, select the **donation** dataset, and then click **Add**.
3. Click **Add a step** Add Step icon on the donation node, and select **Filter**.
4. In Filter, click **Add Filter** Add Filter icon , and then select **DATE\_COMPLETED** from Available data.
5. In DATE\_COMPLETED, under Date Range keep **Range** as the value. In the first field, enter 1/1/2010 as the start date. In the second field enter 12/31/2014 and the end date, and then click inside the dialog.



1. Click **Save**. In Save Data Flow As, enter School Donations, and then click **OK**.

Add Columns

In this section, you create columns by defining expressions.

1. From Data Flow Steps, drag the **Add Columns** to **Add a step** Add a step icon on the Filter node. In **Name**, enter SCH\_STATE. In the expression field, enter the following:

Substring(SCH\_STATEZIP from 1 for 2)

1. Click **Validate**, and then click **Apply**.
2. In Add Columns, click **Column** Add Column icon. In **Name**, enter SCH\_ZIP. In the expression field, enter the following:

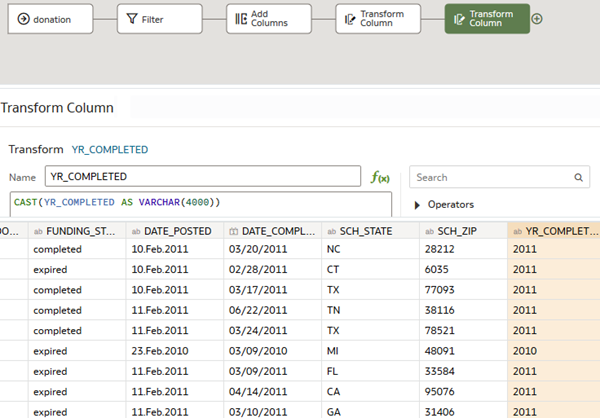
cast(Substring(SCH\_STATEZIP from 4) as int)

1. Click **Validate**, and then click **Apply**.
2. In Add Columns, click **Column** Add step icon. In **Name**, enter YR\_COMPLETED. In the expression field, enter the following:

Year(DATE\_COMPLETED)

1. Click **Validate**, and then click **Apply**.
2. Click **Toggle auto-refresh Data Preview** Toggle auto-refresh Data Preview, and then scroll to view the new columns.
3. Select the **SCH\_ZIP** column, click **Options** Options icon, and then select **Convert to Text**. Select the **YR\_COMPLETED** column, click **Options** Options icon, and then select **Convert to Text**.

The Transform column data flow steps are automatically added to the data flow.

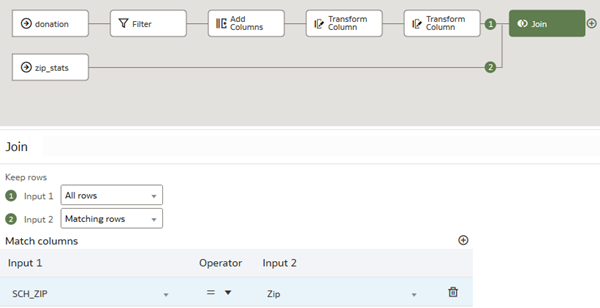


Add a Dataset to the Data Flow

In this section, you add the zip\_stats spreadsheet to the data flow.

1. In Data Flow Steps, double-click **Add Data**. In Add Dataset, click **zip\_stats**, and then click **Add**.
2. In the data flow, click the **Join** step. From Input 1 list, select **All Rows**. Under Match Columns, select **SCH\_ZIP** for the Input 1 value.

The Input 2 value contains the Zip column from the zip\_stats dataset.



Select Columns

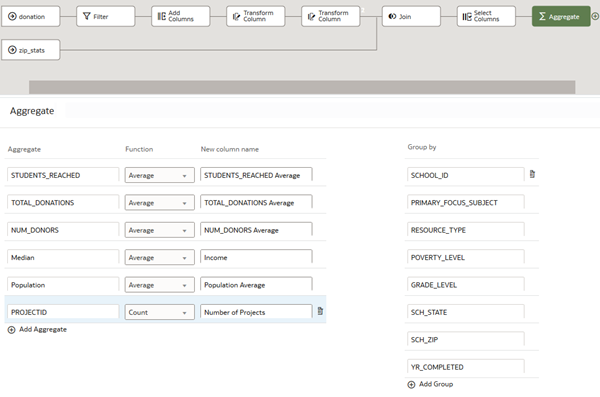
In this section, you select columns to use from the joined datasets.

1. From Data Flow Steps, drag **Select Columns** to **Add a step** Add a step icon on the Join node.

 Under Select Columns, click **Remove All**. Hold down the **Ctrl** key, and select the following columns, and then click **Add Selected**:

* + **PROJECTID**
  + **SCHOOL\_ID**
  + **PRIMARY\_FOCUS\_SUBJECT**
  + **RESOURCE\_TYPE**
  + **POVERTY\_LEVEL**
  + **GRADE\_LEVEL**
  + **STUDENTS\_REACHED**
  + **TOTAL\_DONATIONS**
  + **NUM\_DONORS**
  + **SCH\_STATE**
  + **SCH\_ZIP**
  + **YR COMPLETED**
  + **Median**
  + **Population**

1. From Data Flow Steps, drag **Aggregate** to **Add a step** Add a step icon on the Select Columns node.
2. Under Group by, click **Remove** Delete icon next to the **PROJECTID** row. Click **Add Aggregate**, click in the new field, and then select **PROJECTID** from Available Data.
3. Under Aggregate, from the **Function** list, select **Average** in the following rows:
   * **STUDENTS\_REACHED**
   * **TOTAL\_DONATIONS**
   * **NUM\_DONORS**
   * **Median**
   * **Population**
4. Under Aggregate, click **Add Aggregate**, select **PROJECTID**. From the **Function** list, select **Count**, and then enter Number of Projects in **New column name**. In the **Median** row, enter Income in **New column name**.



Define Aggregation

In this section, you use the Aggregate step to set the functions to use for some columns in the dataset.

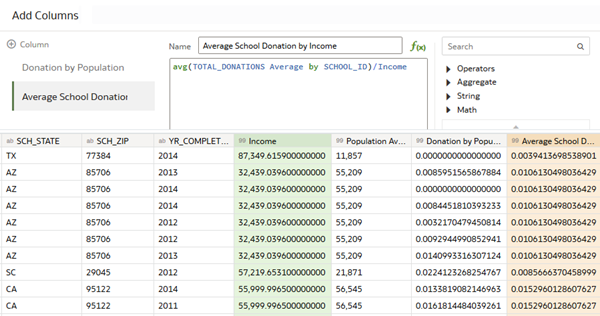
1. From Data Flow Steps, drag **Add Columns** to **Add a step** Add a step icon on the Aggregate node.
2. In **Name**, enter Donation by Population.
3. To represent the average donation amount by the average population in the zip code, in the expression field, enter the following:

TOTAL\_DONATIONS Average/Population Average

1. Click **Validate**. Click **Apply**.
2. In Add Columns, click **Column** Add Column icon. In **Name**, enter Average School Donation by Income.
3. In the expression field, enter the following:

avg(TOTAL\_DONATIONS Average by SCHOOL\_ID)/Income

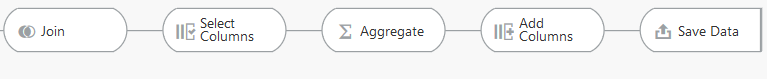
1. Click **Validate**. Click **Apply**.



Save and Run the Data Flow

In this section, you name the dataset that is output as a result of running the data flow, and examine the columns in the new dataset. You use the Donations by School dataset in the next tutorial.

1. From Data Flow Steps, drag **Save Dataset** to the last Add a step Add a step icon node on the data flow.
2. In Save Dataset, enter Donations by School. From Save data to, select **Dataset Storage** to save the data in Oracle Analytics.
3. Click **Save**.



1. Click **Run Data Flow** Run Data Flow icon.
2. After the data flow run completes successfully, click **Go back** Back icon.
3. On the Home page, select the **Donations by School** dataset, click the **Actions Menu** Actions Menu icon, and then select **Inspect**.
4. In Donations by School Dataset, click **Data Elements** to view the columns in the dataset.

