

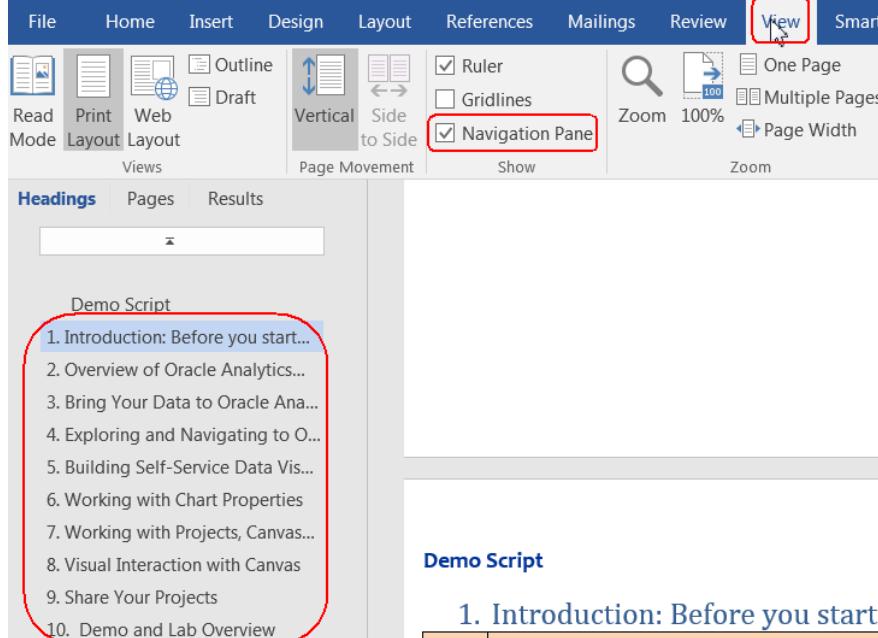
Assignment: Data Flow Deep Dive with Oracle Analytics

Table of Contents

Note	3
Demo Script	4
1. Project Overview: University Donation Analysis	4
2. Use Case and Key Features of Data Flow	6
3. Create Aggregation Attribute as a Data Flow Step.....	8
4. Run and Manage Data Flow	22
5. Create Multiple Target Data Set Using a Branch in a Data Flow	31
6. Advanced Properties of Data Flow	38
7. Working with Incremental Data with Data Flow	46
8. Create Sentiment Analysis and Enrich Data Set with Data Flow Step	67

Note

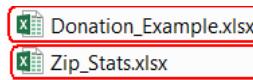
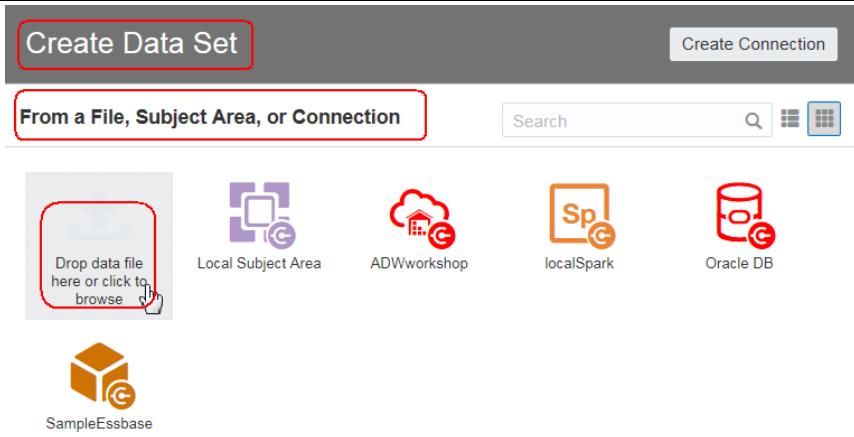
We recommend to enable **Navigation Pane (View Tab)** in your MS Office Word, for an optimal use of this Lab document.

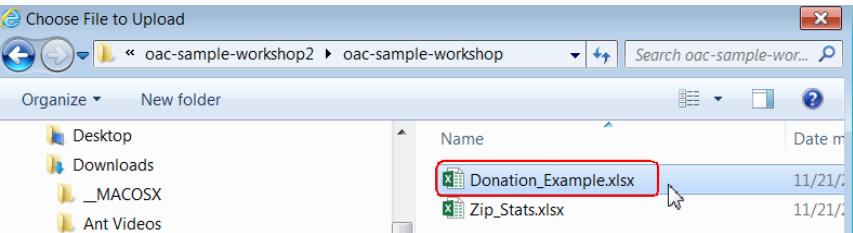
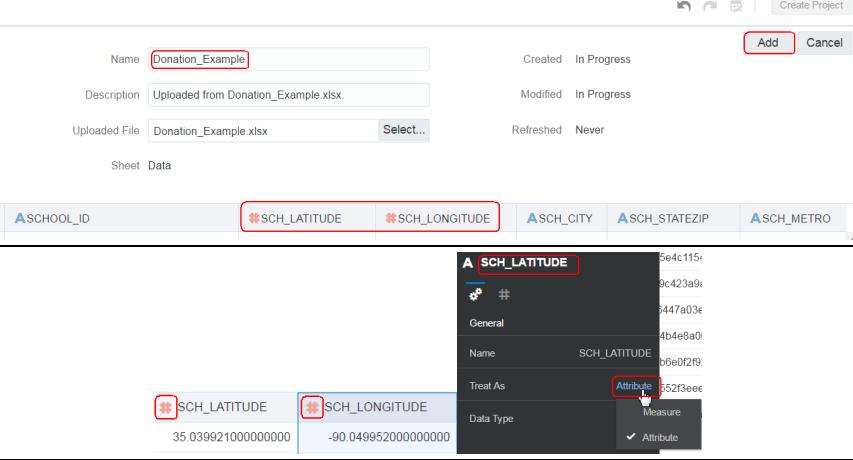
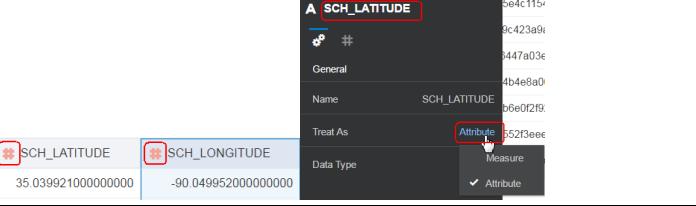


Demo Script

1. Project Overview: University Donation Analysis

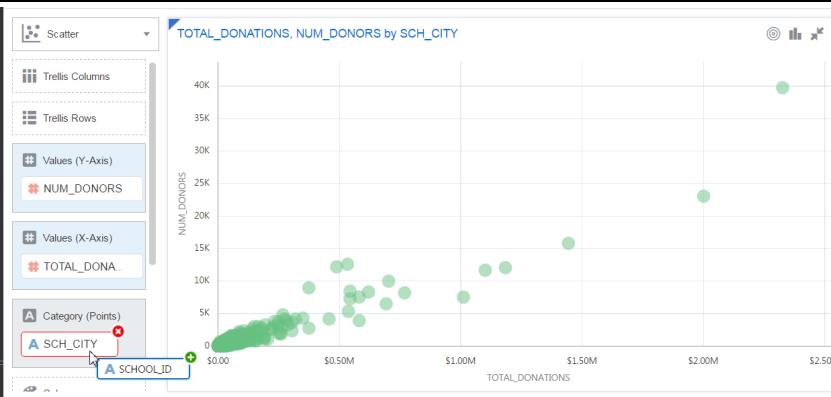
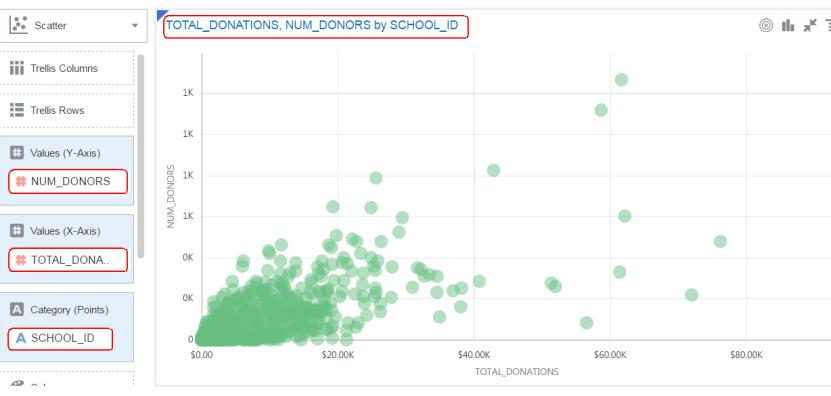
- Navigate to the .xlsx file for this workshop

Step	View	Click stream	Talk stream
1.			<p>What do you need:</p> <ul style="list-style-type: none"> - This project uses these two xlsx files to showcase dataflow capabilities <ul style="list-style-type: none"> o The Donation_Example dataset has 150k rows of Schools Projects Donations information spread across 5 years o The Zip_Stats dataset has ~32k rows of Income and Population information at US ZIPcodes level
2.			<p>Data is provided by www.donorschoose.org online charity, which seek US public school project funding by donors.</p> <p>DonorsChoose.org data set it is often used in data science challenges with analysis and visualizations toward questions as: why some projects get funding and some not / does proximity to school change donation behavior and so on</p>
3.			<p>Let's start by upload the Donation_Example.xls file and create a dataset. Give it a bit of time to upload the dataset...</p> <p>Note: we have been created data sets on previous lectures, therefore we are not repeating all the detailed steps, but rather you get the relevant print screens</p>

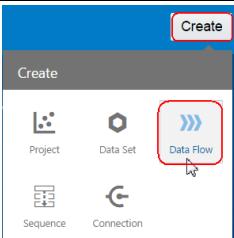
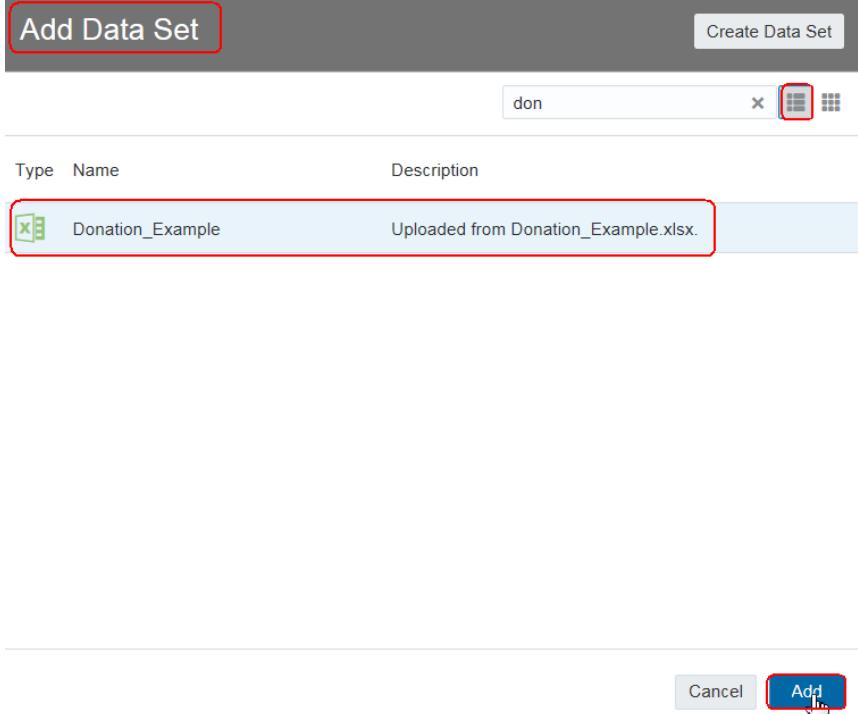
Step	View	Click stream	Talk stream
4.			
5.			
6.		-	Latitude and Longitude were marked as metrics (numbers). Let's set them as attributes
7.		- Click Add	
8.		-	Save your data set as Donation Examples

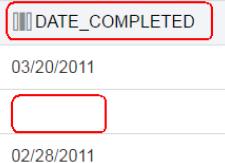
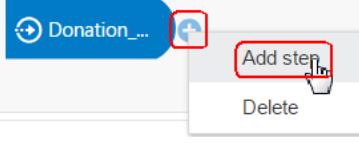
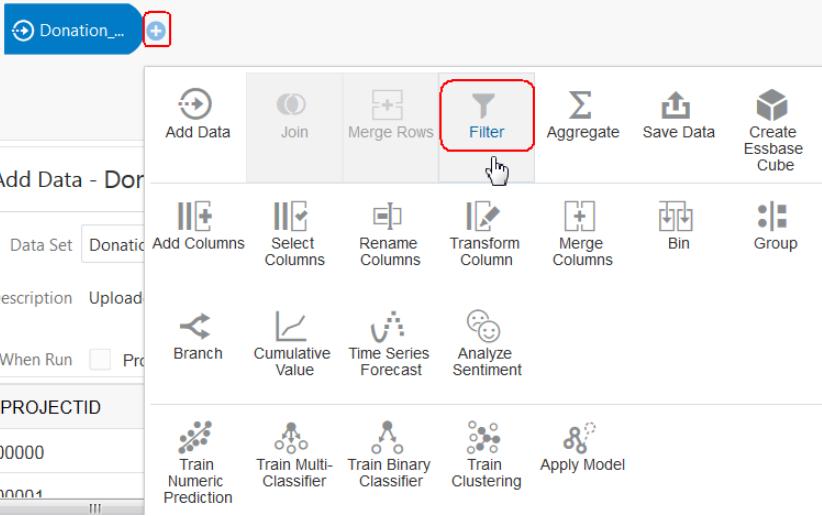
2. Use Case and Key Features of Data Flow

Step	View	Click stream	Talk stream
1.			<p>Run simple analysis on the data-set</p> <p>While the dataset is being profiled, we can start analyzing it</p>
2.			<p>Let's look at Total Donations amount by Year first</p> <p>TOTAL_DONATIONS by DATE_COMPLETED_YEAR</p>
3.			<p>Now, a scatter of Total Donations x Number of Donors per City</p> <p>TOTAL_DONATIONS NUM_DONORS by SCH_CITY</p> <p>In our dataset, Donations and Donors are highest across LA, followed by Chicago</p>

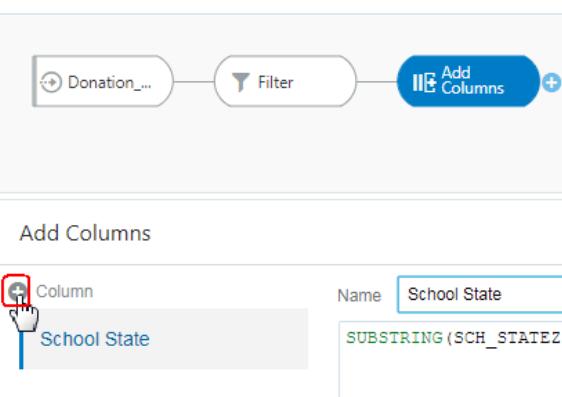
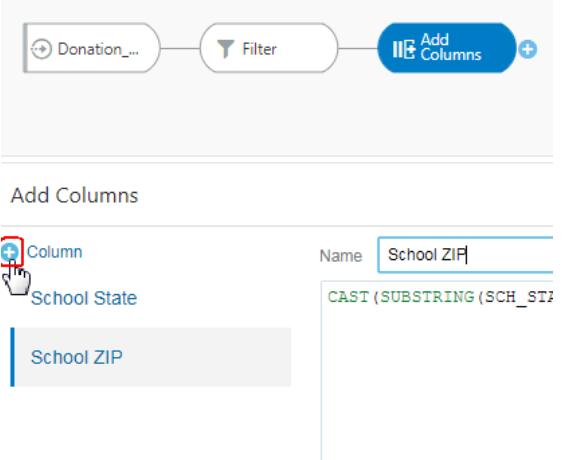
Step	View	Click stream	Talk stream
4.	 <p>TOTAL_DONATIONS, NUM_DONORS by SCH_CITY</p> <p>Scatter</p> <ul style="list-style-type: none"> Trellis Columns Trellis Rows Values (Y-Axis) NUM_DONORS Values (X-Axis) TOTAL_DONATIONS Category (Points) SCH_CITY A SCHOOL_ID 		<p>We can do the same by Schools...</p> <p>TOTAL_DONATIONS NUM_DONORS by SCHOOL_ID</p>
5.	 <p>TOTAL_DONATIONS, NUM_DONORS by SCHOOL_ID</p> <p>Scatter</p> <ul style="list-style-type: none"> Trellis Columns Trellis Rows Values (Y-Axis) NUM_DONORS Values (X-Axis) TOTAL_DONATIONS Category (Points) SCHOOL_ID 		
6.			<p>This is good initial insight, but we want to dig deeper and look at this in the perspective of metrics like Population, Income which are part of our other data set (ZIP Stats).</p> <p>To achieve we could add the zip dataset to this project, or, we could create a dataflow to derive a new merged dataset that will help us compare different schools by these various metrics. Let's do that</p>

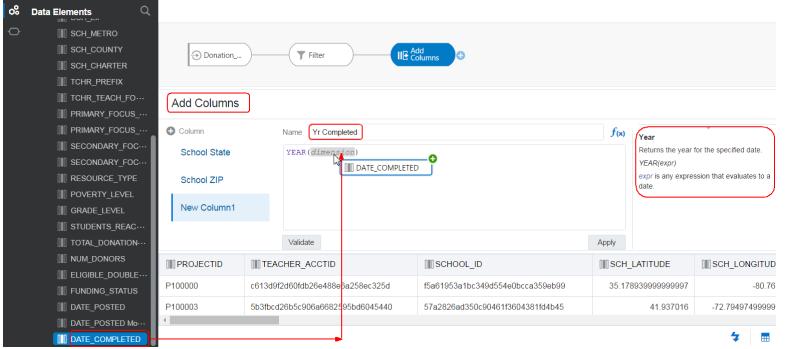
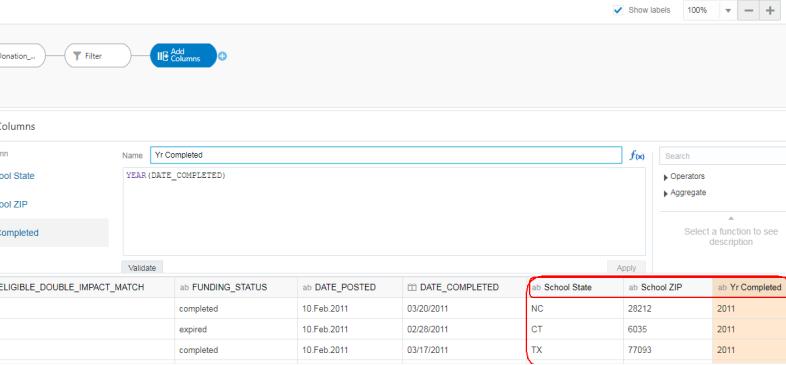
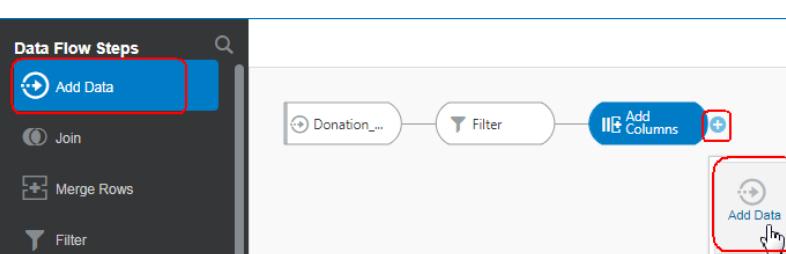
3. Create Aggregation Attribute as a Data Flow Step

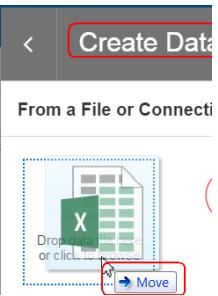
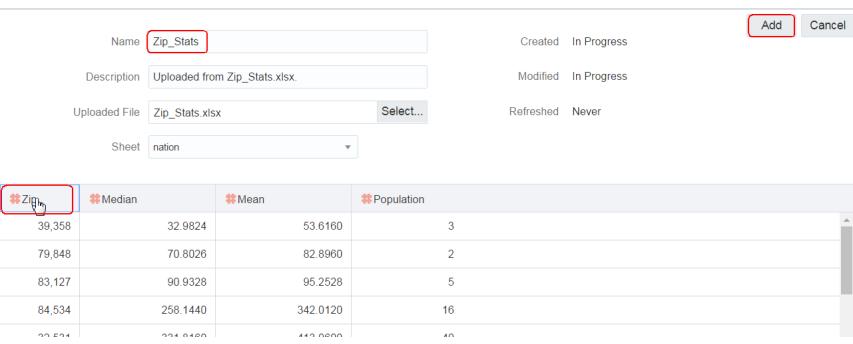
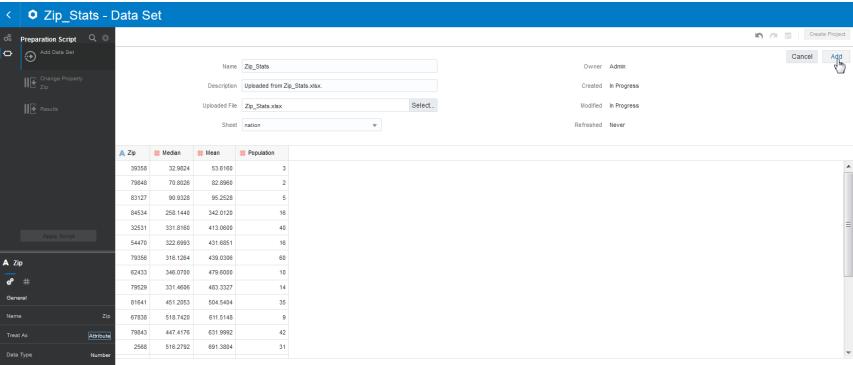
Step	View	Click stream	Talk stream
1.			Let's create your first Data Flow
1.			A Data Flow is a user-friendly lightweight transformation tool. It provides a clean and powerful interface to a business user to cleanse and transform your data.
2.		Click Create > Data Flow	
3.			Add Data Set Donation_Examples

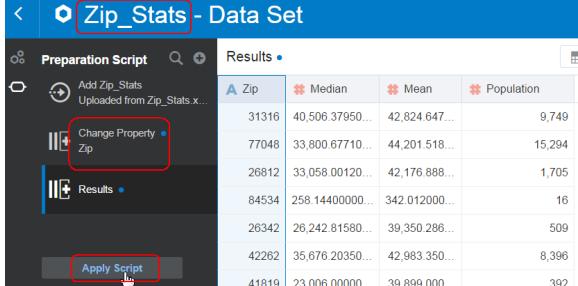
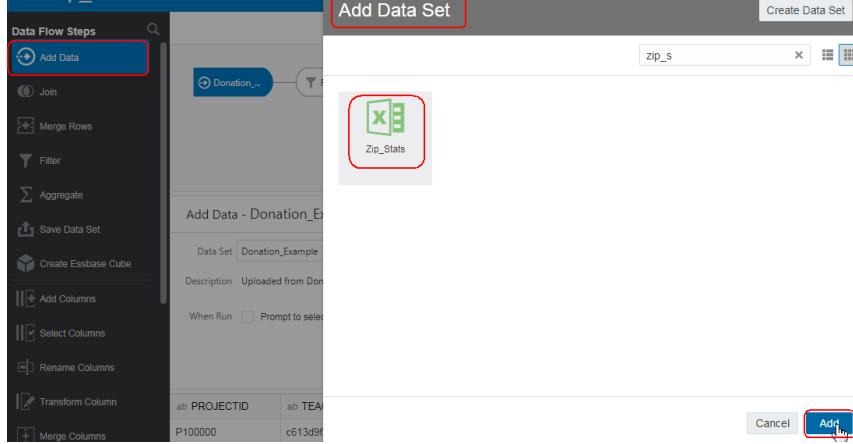
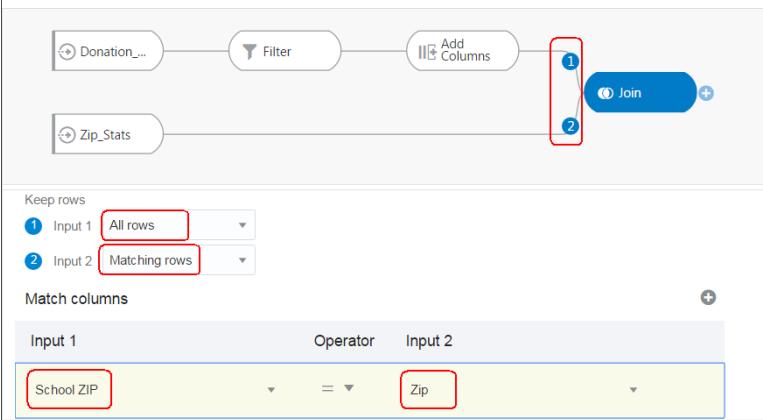
Step	View	Click stream	Talk stream
4.			<p>Data Flows allow us to create sequential transformations of our data, and save the results in a new dataset.</p> <p>There are a number of steps to enable reasonably complex transformations on the source data, with a single drag/drop experience.</p>
5.			<p>There are some ongoing projects whose completion date is null.</p>
6.		<ul style="list-style-type: none"> - Right click on  - Select Add step 	<p>Let's first Filter these records out.</p> <p>In a filter step, you can either drag/drop a column and filter the values or use the expression filter option for more complex filtering.</p>
7.		<ul style="list-style-type: none"> - Click on  	<p>Notice that steps as Join and Merge Rows are greyed out, as there are not possible to choose at this point in time</p>

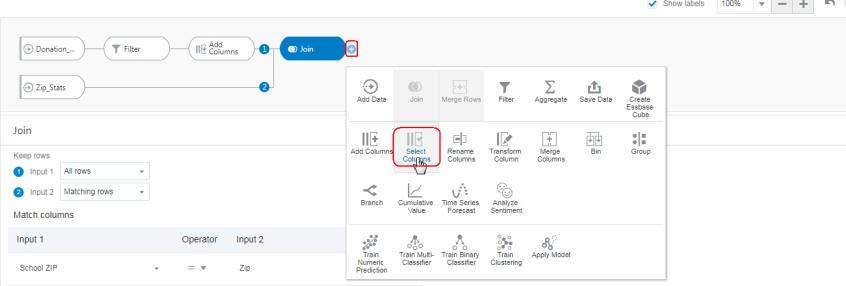
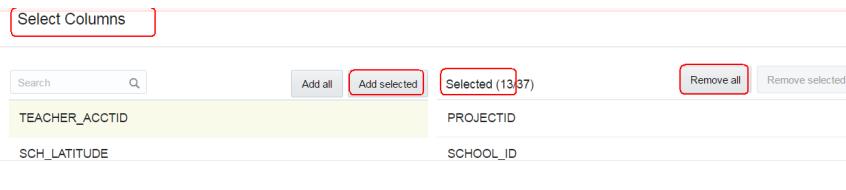
Step	View	Click stream	Talk stream
8.		<ul style="list-style-type: none"> - Click on - Add Expression Filter 	Let's use an expression filter to remove the NULL values in the data column
9.		<p>DATE_COMPLETED IS NOT NULL</p> <ul style="list-style-type: none"> - Click Validate - Click Apply 	The autocomplete feature suggests column names and functions making it easier to enter your expressions
10.		<ul style="list-style-type: none"> - Right click on - Select Add step - Click Add Column 	The dataset has school state and zip code as part of the same column. Let's use the Add Columns step to extract State and Zip information and also extract the Year from the Date Completion column
11.		<p>Type Name: School State SUBSTRING(SCH_STATEZIP FROM 1 FOR 2)</p> <ul style="list-style-type: none"> - Click Validate - Click Apply 	School State

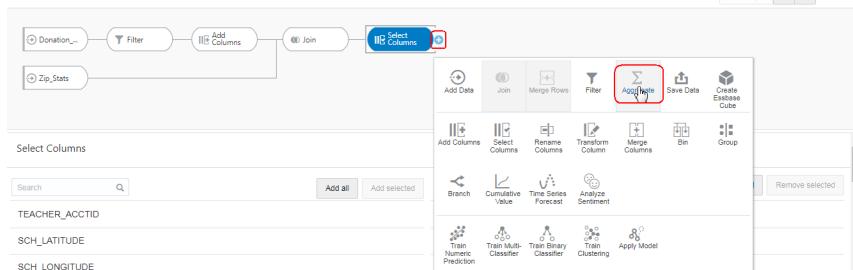
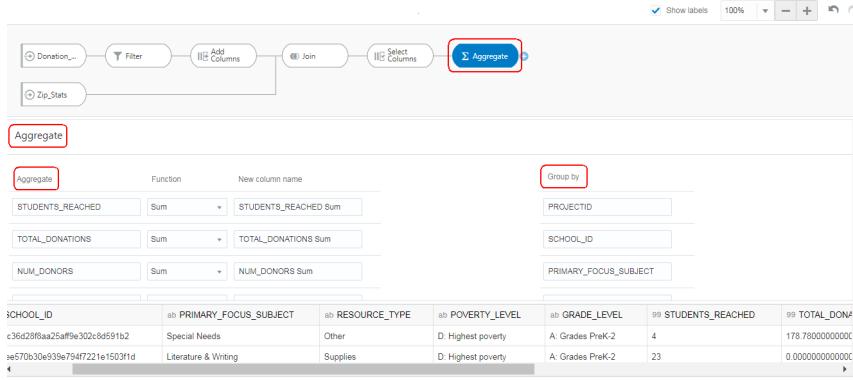
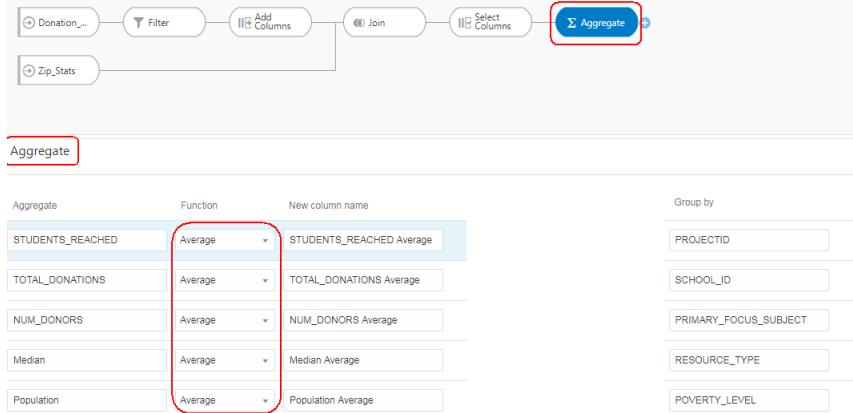
Step	View	Click stream	Talk stream												
12.	 <p>Add Columns</p> <p>Column Name: School State SUBSTRING(SCH_STATE1)</p>	<ul style="list-style-type: none"> - Click on  	Add a New Column												
13.	<p>Add Columns</p> <table border="1"> <tr> <td>Column</td> <td>Name: School ZIP</td> <td></td> </tr> <tr> <td>School State</td> <td>CAST(SUBSTRING(SCH_STATEZIP FROM 4) AS int)</td> <td></td> </tr> <tr> <td>School ZIP</td> <td colspan="2"> Calculation validated</td> </tr> <tr> <td></td> <td>Validate</td> <td></td> </tr> </table>	Column	Name: School ZIP		School State	CAST(SUBSTRING(SCH_STATEZIP FROM 4) AS int)		School ZIP	 Calculation validated			Validate		<p>Type Name: School ZIP CAST(SUBSTRING(SCH_STATEZIP FROM 4) AS int)</p> <ul style="list-style-type: none"> - Click Validate - Click Apply 	<p>School ZIP</p> <p>Let's cast the zipcode as an integer so it can be used to join to the ZIP Stats dataset in the next step</p>
Column	Name: School ZIP														
School State	CAST(SUBSTRING(SCH_STATEZIP FROM 4) AS int)														
School ZIP	 Calculation validated														
	Validate														
14.	 <p>Add Columns</p> <p>Column Name: School ZIP CAST(SUBSTRING(SCH_STATEZIP FROM 4) AS int)</p>	<ul style="list-style-type: none"> - Click on  	Add a New Column												

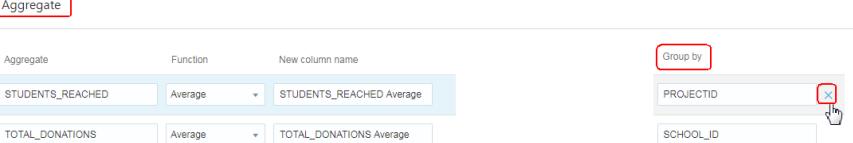
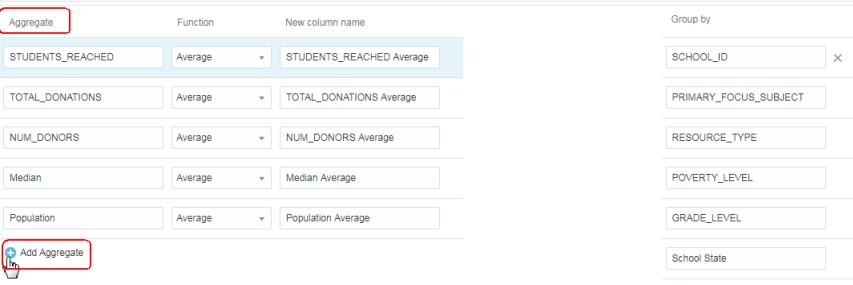
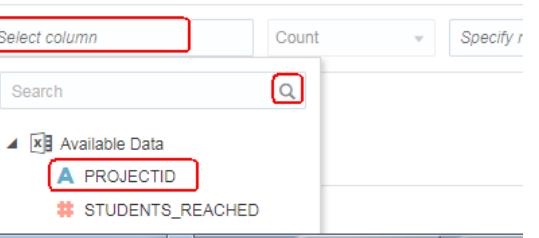
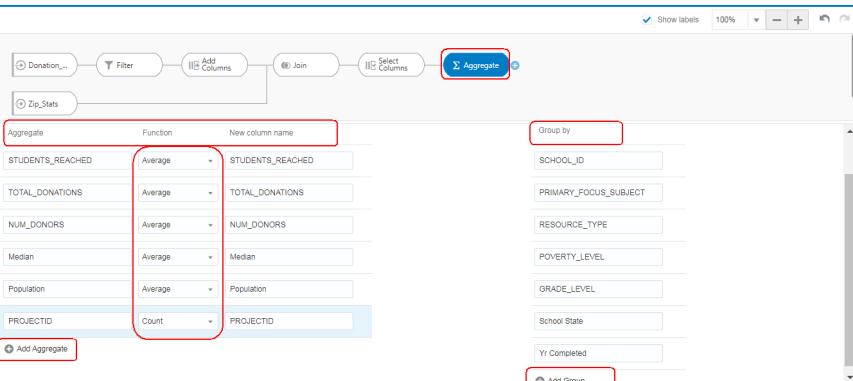
Step	View	Click stream	Talk stream
15.		<p>Type Name: Yr Completed YEAR(DATE_COMPLETED)</p> <ul style="list-style-type: none"> - Click Validate - Click Apply 	<p>Yr Completed</p> <p>We can either type in an expression or choose from an exhaustive list of functions that are available)</p> <p>Every function has a brief description as well as sample syntax</p>
16.			<p>The three new columns we just added are available as part of the data preview.</p> <p>Data preview is available at every step within a data flow.</p>
17.			<p>Next, let us add the dataset based on the Zip_Stats.xlsx file to the dataflow. We will create it on the fly.</p> <p>With the second Data Set you grab information about Income and Population</p>
18.		<ul style="list-style-type: none"> - Right click on  - Select Add step - Select Add Data - Click Create Data Set 	<p>> Add Data Set Create new Data Set</p>

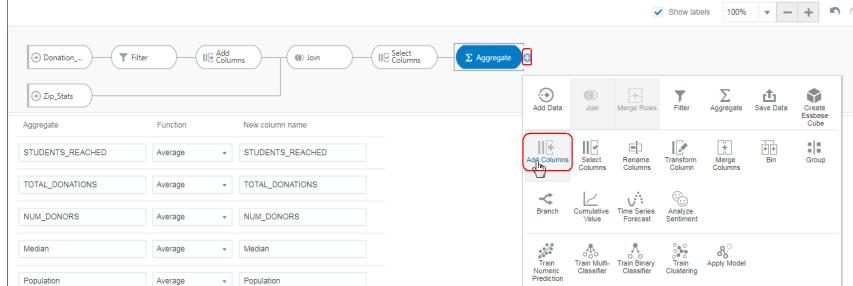
Step	View	Click stream	Talk stream
19.		- Drag and drop Zip_Stats.xlsx	
20.		- Click Add	<p>Numeric columns are set as Measures by default. Let's set the Zip as an attribute.</p> <p>Note: If the change doesn't work on Data Flow context, then you create your data set as a separate step and change (ZIP > Attribute)</p>
21.		-	<p>Hint: if you work on Oracle Analytics, you just open a new browser and create the ZIP_STATS data set</p>

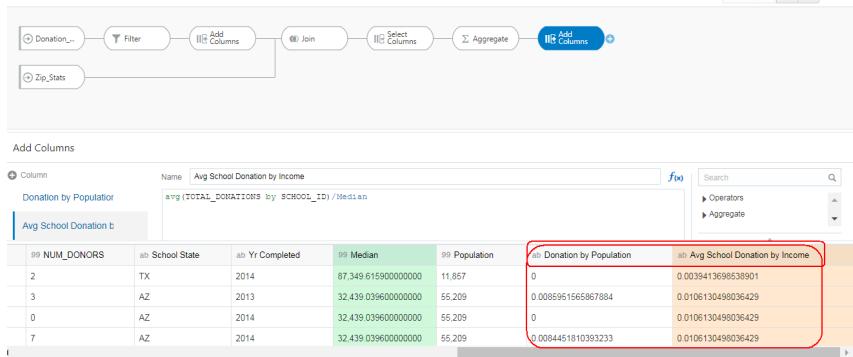
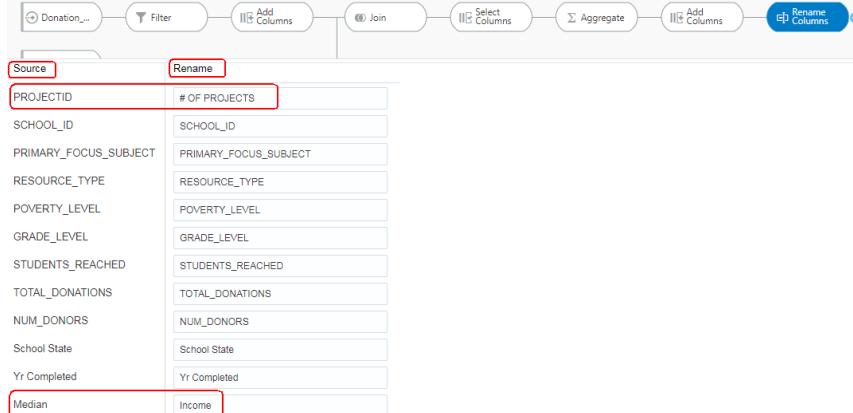
Step	View	Click stream	Talk stream
22.	 <p>The screenshot shows the 'Zip_Stats - Data Set' view. A preparation script named 'Add_Zip_Stats' is displayed. A step titled 'Change Property Zip' is highlighted with a red box. At the bottom, the 'Apply Script' button is also highlighted with a red box.</p>		
23.	 <p>The screenshot shows the 'Add Data Set' dialog box. It lists 'Add Data' as the selected step. A dataset named 'zip_s' is selected from a dropdown menu. The 'Add' button at the bottom right is highlighted with a red box.</p>		
24.	 <p>The screenshot shows the Data Flow Designer interface. A 'Join' step is being added between two datasets: 'Donation...' and 'Zip_Stats'. The 'Keep rows' section shows 'Input 1: All rows' and 'Input 2: Matching rows' both set to 'All rows'. The 'Match columns' section shows 'Input 1: School ZIP' and 'Input 2: Zip' connected by an equals sign. The 'Join' step itself is highlighted with a red box.</p>	<p>Input 1: All rows (School Zip) Input 2: Matching rows (Zip)</p>	<p>Adding a dataset in the dataflow automatically created a Join step in the flow.</p> <p>We can set an outer join behavior by choosing to have All Rows from an input. Let's choose All Rows from our Donations dataset as there may be zip codes which don't exist in the Zip Stats dataset.</p>

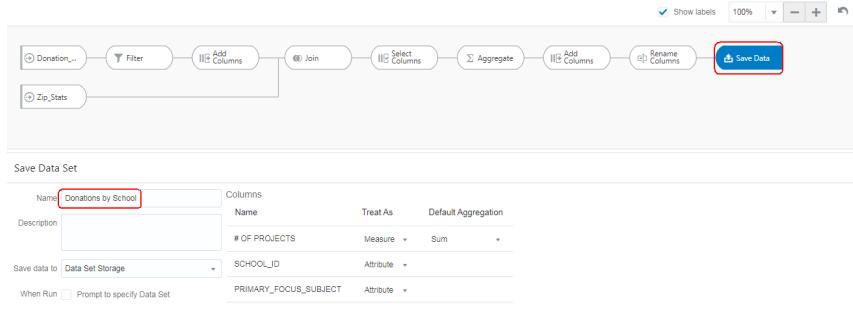
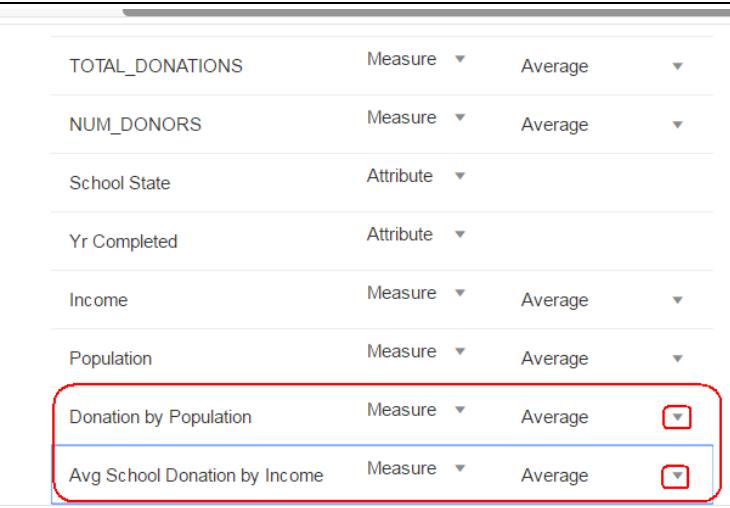
Step	View	Click stream	Talk stream
25.			The combined dataset now holds several columns . Let us keep only the ones we are interested in .
26.		<ul style="list-style-type: none"> - Right click on  Select Columns 	<p>Let's use Select Columns step and pick only a few relevant Attributes and Measures</p>
27.		<ul style="list-style-type: none"> - Click Remove All - Keep Ctrl pressed and select: <ul style="list-style-type: none"> ○ PROJECTID ○ SCHOOL_ID ○ PRIMARY_FOCUS SUBJECT ○ RESOURCE_TYPE ○ POVERTY_LEVEL ○ GRACE_LEVEL ○ STUDENTS_REACHED ○ TOTAL_DONATIONS ○ NUM_DONORS ○ School State ○ Yr Completed ○ Median ○ Population - Click Add Selected 	<p>Data preview can be switched on/off</p> 

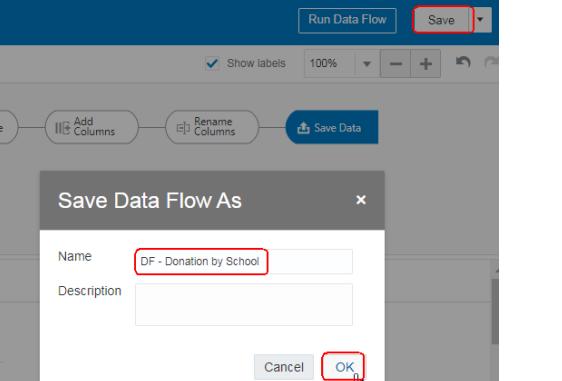
Step	View	Click stream	Talk stream
28.		<ul style="list-style-type: none"> - Right click on  + - Select Aggregate 	Next, let's aggregate our dataset using the Aggregate step > + Aggregate
29.			Metrics are on the left and Attributes on the right side
30.		<p>STUDENTS_REACHED > Average TOTAL_DONATIONS > Average NUM_DONORS > Average Population > Average Median > Average</p>	<p>You just have to set aggregation rules for the metrics, or flag them as non-aggregated</p> <p>Use the print screen as a guide</p> <p>Hint: toggle off Preview Data  to have a better look on the rows</p>

Step	View	Click stream	Talk stream
31.		- Hover over PROJECTID and click on 	Note: to move a data element from measure to attribute, hover the mouse over the data element, click on the X icon 
32.		- Click on  - Go to lower left and Add (Aggregate) PROJECTID	
33.		- Click on Select Column and select PROJECTID	
34.			Use the print screen as guidance. You may want the New Column Name to have the same name as original aggregated columns
35.			Now that our dataset is aggregated at the level we need, let's create a couple of calculations to look at Donations at the school level

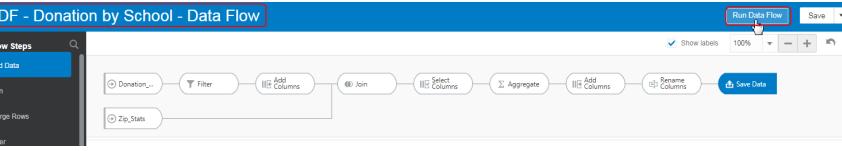
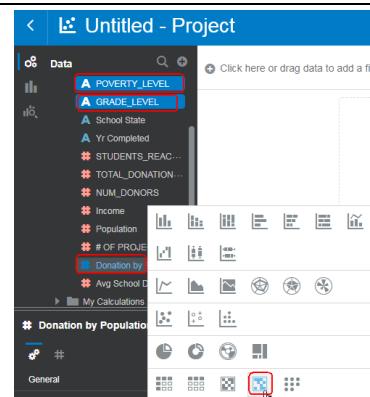
Step	View	Click stream	Talk stream
36.		<ul style="list-style-type: none"> - Right click on  - Select Add step - Click Add Column 	
37.		<p>Type Name: Donation by Population TOTAL_DONATIONS/Population</p> <ul style="list-style-type: none"> - Click Validate - Click Apply 	Avg Donation Amounts by Avg Zip Code Population
38.		<ul style="list-style-type: none"> - Click on  <p>Type Name: Avg School Donation by Income avg(TOTAL_DONATIONS by SCHOOL_ID) / Median</p> <ul style="list-style-type: none"> - Click Validate - Click Apply 	Ratio of Avg Donation Amounts by Median Income in the zip area
39.		<ul style="list-style-type: none"> - Click on  	Toggle on Preview Data

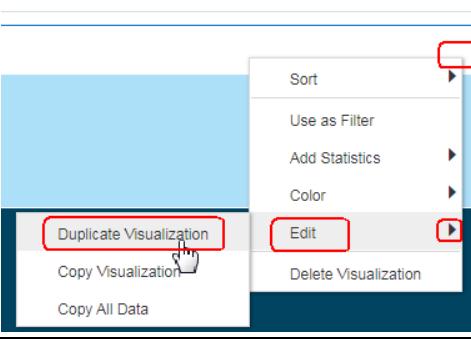
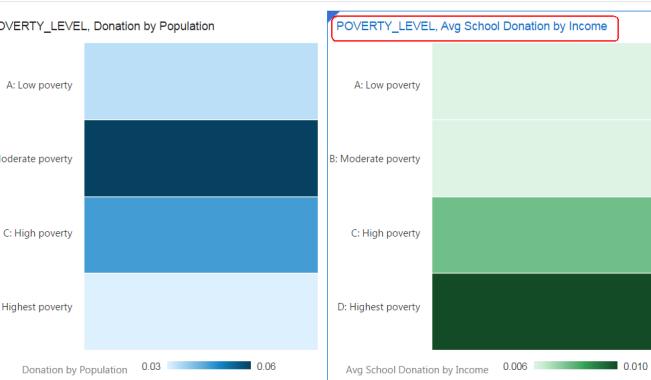
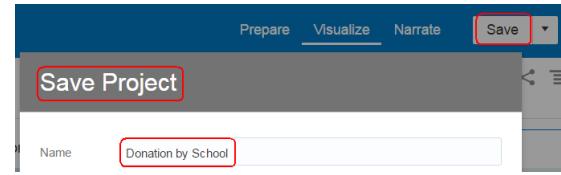
Step	View	Click stream	Talk stream
40.	 <p>The screenshot shows the Oracle Analytics Data Flow interface. In the preview pane, there is a table with columns: # NUM_DONORS, ab School State, ab Yr Completed, # Median, # Population, ab Donation by Population, and ab Avg School Donation by Income. One row is selected and highlighted with a red box. The 'Add Columns' step is visible at the bottom of the flow.</p>		New columns are visible in the preview pane
41.			Let's use the Rename Columns step to rename some of the columns of our dataset
42.	 <p>The screenshot shows the Oracle Analytics Data Flow interface. The context menu for the 'Add' step in the flow is open, showing various options like Add Data, Join, Merge Rows, Filter, Aggregate, Save Data, Create Essbase Cube, Add Columns, Select Columns, Rename Columns, Transform Column, Merge Columns, Bin, and Group. The 'Rename Columns' option is highlighted with a red box.</p>	<ul style="list-style-type: none"> - Right click on  - Select Add step - Click Rename Columns 	
43.	 <p>The screenshot shows the Oracle Analytics Data Flow interface. The 'Rename' step in the flow has several columns highlighted with red boxes: PROJECTID, # OF PROJECTS, SCHOOL_ID, PRIMARY_FOCUS SUBJECT, RESOURCE_TYPE, POVERTY_LEVEL, GRADE_LEVEL, STUDENTS_REACHED, TOTAL_DONATIONS, NUM_DONORS, School State, Yr Completed, Median, and Income. The 'Rename' step itself is also highlighted with a red box.</p>	PROJECTID > # OF PROJECTS Median > Income	Rename Columns

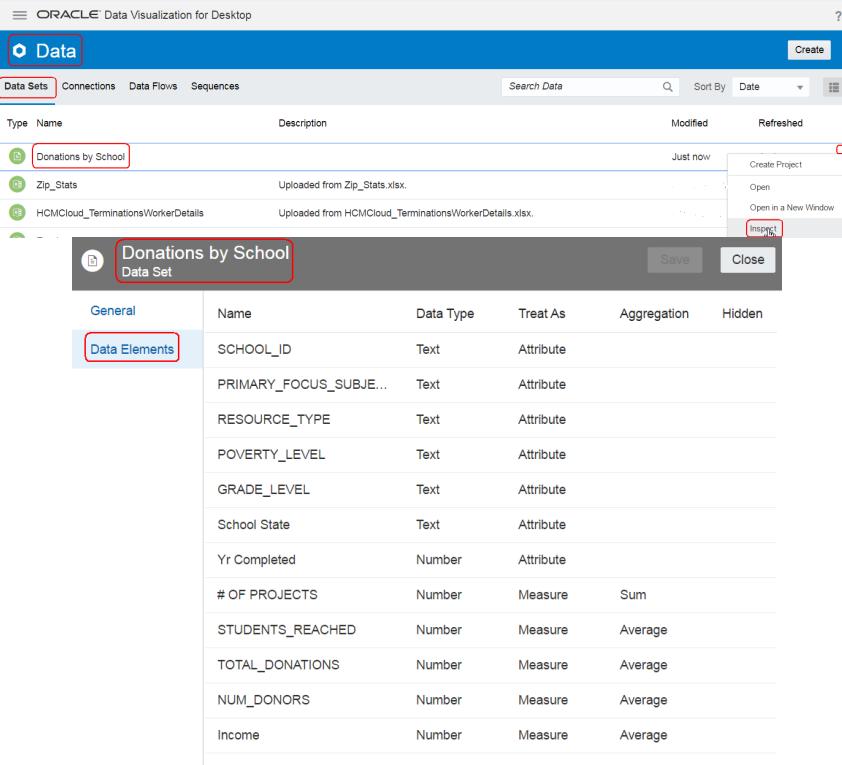
Step	View	Click stream	Talk stream
44.		<ul style="list-style-type: none"> - Right click on  - Select Add step - Click Save Data 	<p>Let's now save the dataset</p> <p>> + Save Data</p> <p>Save Data Set</p>
45.		<p>Name: Donations by School</p> <p>Save data to: Data Set Storage</p>	<p>Save data to: Data Set Storage or Database Connection</p> <p>Output dataset can either be stored to the Data Set Storage (inside Oracle Analytics) or as a database table.</p> <p>In case of a database, we need to select a connection to the database.</p> <p>You can have a Prompt on the output dataset name and specify the name at run time when the dataflow is executed.</p>
46.		<p>Donation by Population > Average Avg School Donation by Income > Average</p>	<p>You can set the default aggregation rules of each measures on the output dataset</p>

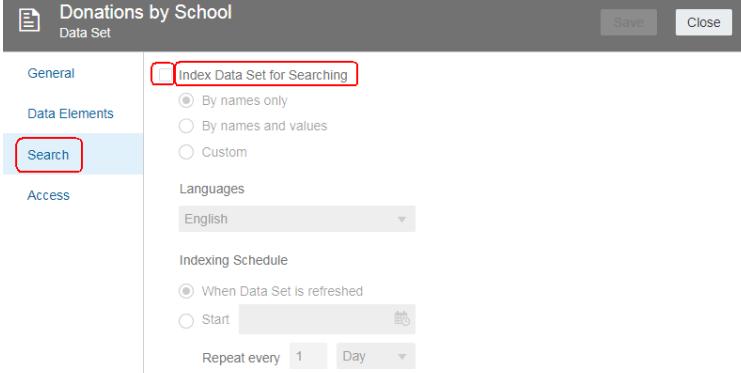
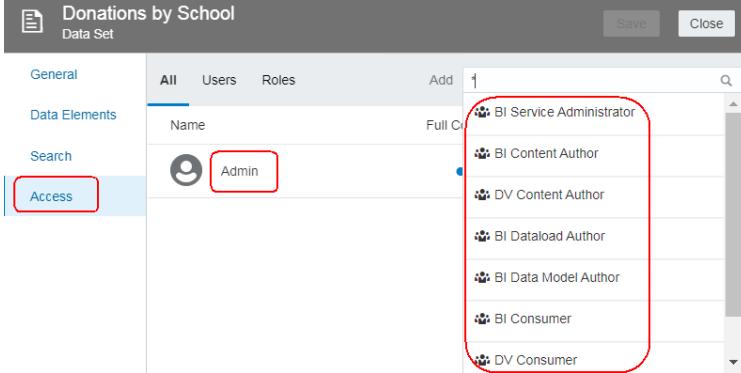
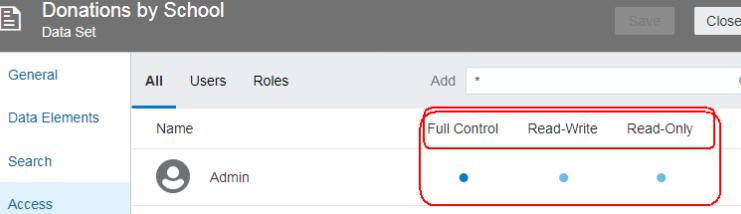
Step	View	Click stream	Talk stream
47.	 <p>Save Data Flow As</p> <p>Name: DF - Donation by School</p> <p>Description:</p> <p>Cancel OK</p> <p>Data Flow "DF - Donations by School" complete.</p>	<ul style="list-style-type: none"> - Click Save - Name: DF - Donation by School - Click Ok <p>Run Data Flow Run Data Flow</p>	<p>Now let's Save the Data Flow</p>

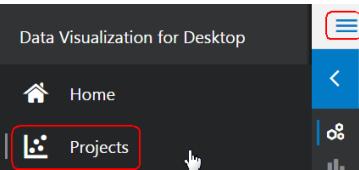
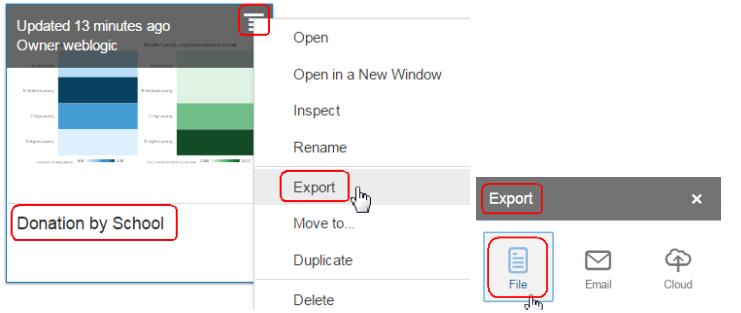
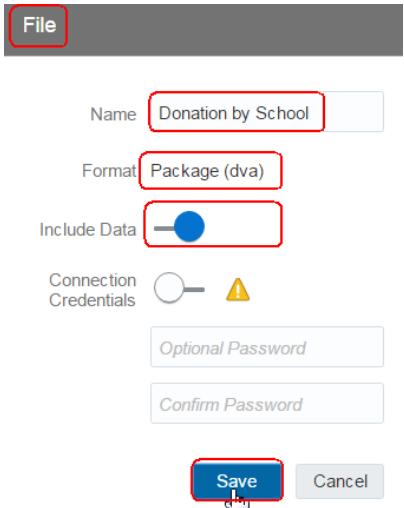
4. Run and Manage Data Flow

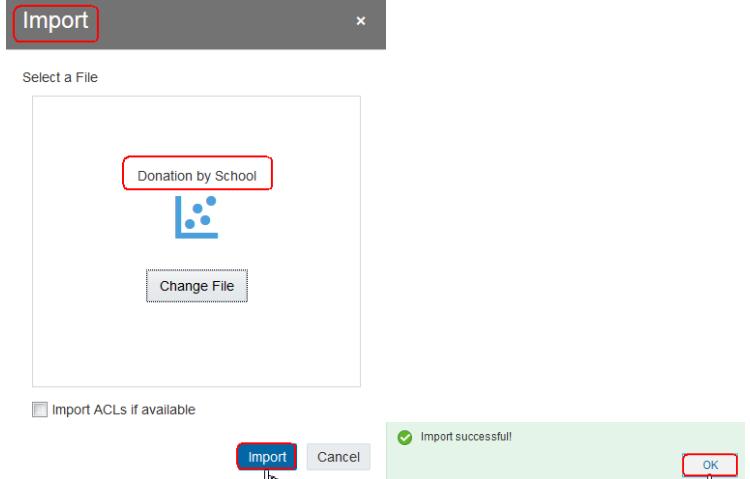
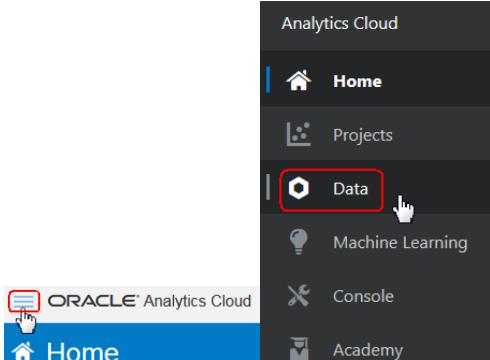
2.		Run Data Flow	Now let's Run the Data Flow
3.			Let's analyze the output dataset (Donation by School Data Set.)
4.		- Go to Actions Menu	Create Project.
5.		<ul style="list-style-type: none"> - Select: <ul style="list-style-type: none"> o POVERTY_LEVEL o GRADE_LEVEL o Donation by Population - Right Click > Pick Visualization > Grid Heat Map 	
6.		<ul style="list-style-type: none"> - Remove GRADE_LEVEL 	<p>POVERTY_LEVEL Donation by Population</p> <p>Highest average Donation by Population is for Grades 3-5, "moderate poverty" school areas</p> <p>Note: "poverty level" of the school requesting the money, defined by the DonorsChoose.org charity according to the percentage of students qualifying for a free or discounted lunch</p>

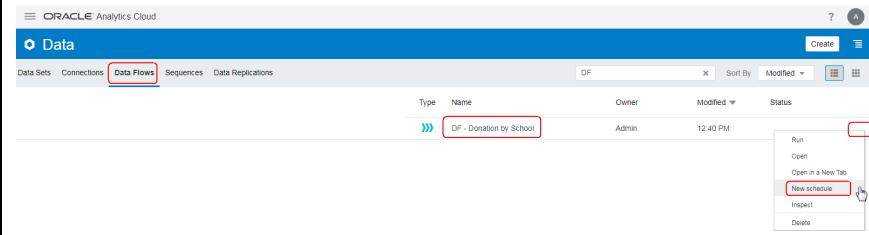
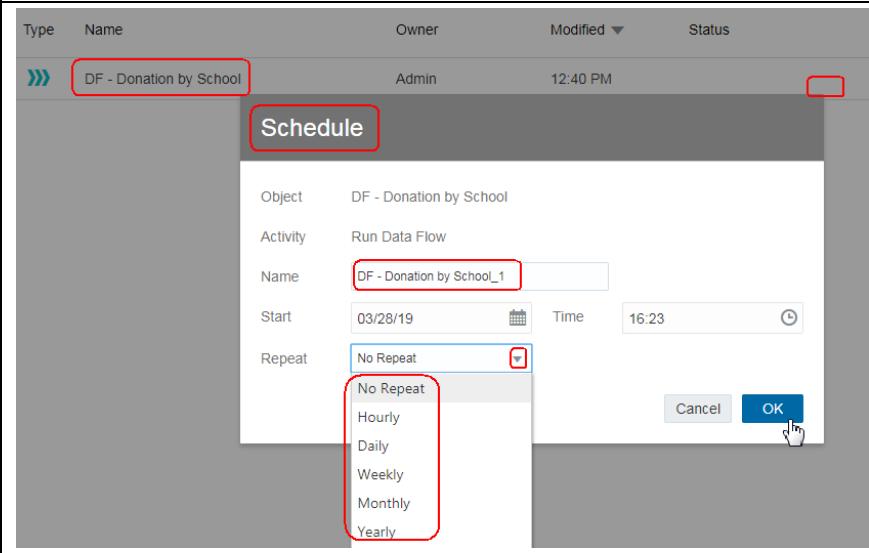
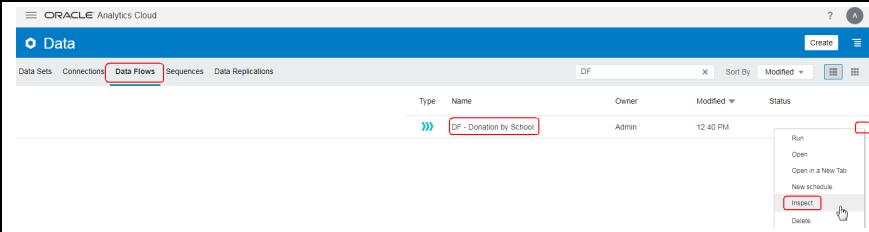
7.		<ul style="list-style-type: none"> - Click on Menu  Edit > Duplicate Visualization 	
8.		<ul style="list-style-type: none"> - Select Avg School Donation by Income and replace Donation by Population 	<p>POVERTY_LEVEL Avg School Donation by Income</p> <p>Highest Avg School Donation shows a different pattern, so does Total Donations metric</p>
9.		<ul style="list-style-type: none"> - Click Save - Name: Donation by School - Click Save 	<p>Save the Project as Donation by School</p>
10.			<p>Now let's Manage the Data Flow</p>

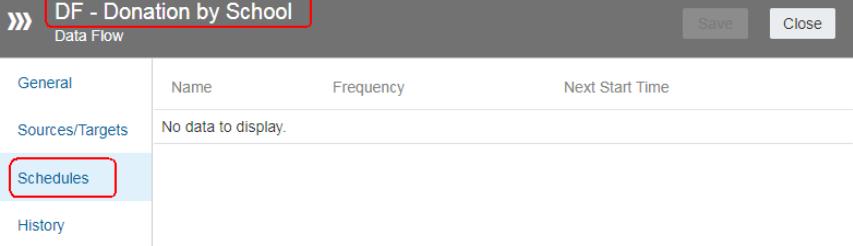
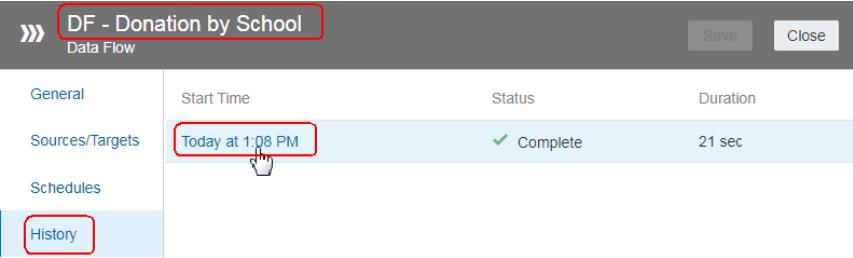
11.		<ul style="list-style-type: none"> - Go to Home Page - Select Data > Data Set - Select Donation by School - Click Inspect <p>Let's inspect the output dataset (Donation by School) that was created by the dataflow execution</p> <p>Aggregations that we set in the Save Data step have been applied on the output dataset</p>
12.		<p>Optional: The same Inspect dataset in Oracle Analytics brings two additional functionalities, Search and Access</p>

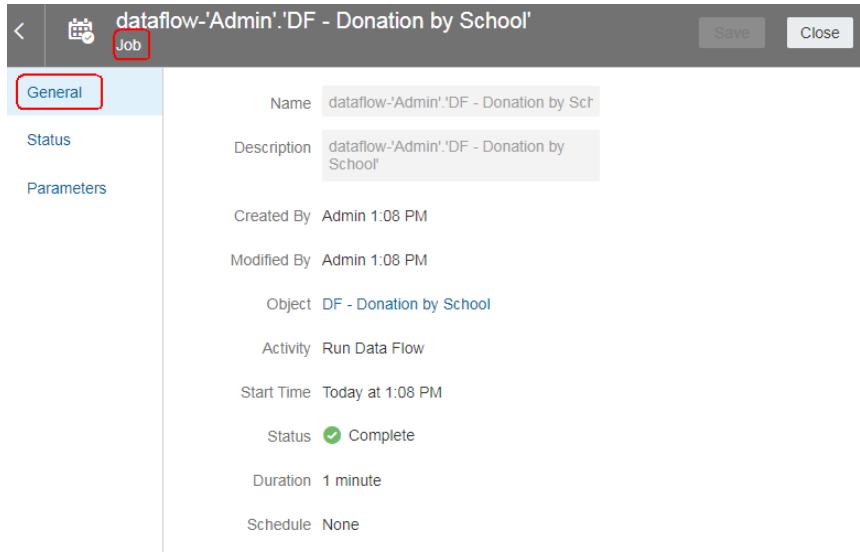
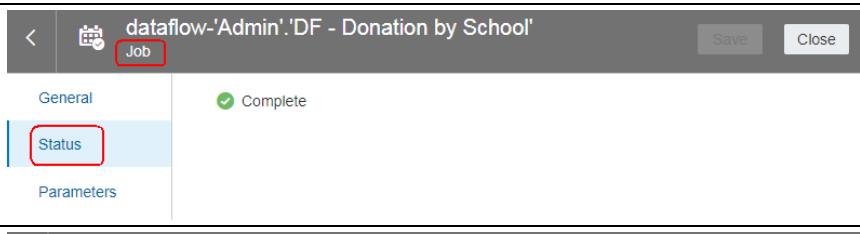
13.		-	<p>Search: you can enable the dataset to be available for search in Day by Day mobile application</p>
14.		-	<p>Access: you can grant access to Users or Roles level</p>
15.		-	<p>Access: you can grant Full Control or Read-Write or just Read-Only for the data set to specific Users or Roles</p>
16.			<p>In addition to ad hoc execution like we just did, a data flow can also be scheduled to execute at a particular date/time with a repeat frequency</p> <p>Schedule it's a feature of in Analytics Cloud, so we are going to export Donation by School project to Analytics Cloud</p>

17.			Let's go to Projects and Export Donation by School
18.		<ul style="list-style-type: none"> - Click on Actions Menu  - Select Export - Select File 	We are going to use .dva package file
19.			Use the print screen as guidance
20.		<ul style="list-style-type: none"> - Click on Page Menu  - Select Import Project 	<p>Optional (if you have access to an Oracle Analytics instance)</p> <p>Go to Oracle Analytics and Import Project</p>

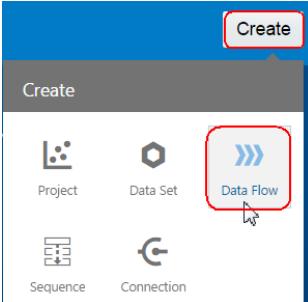
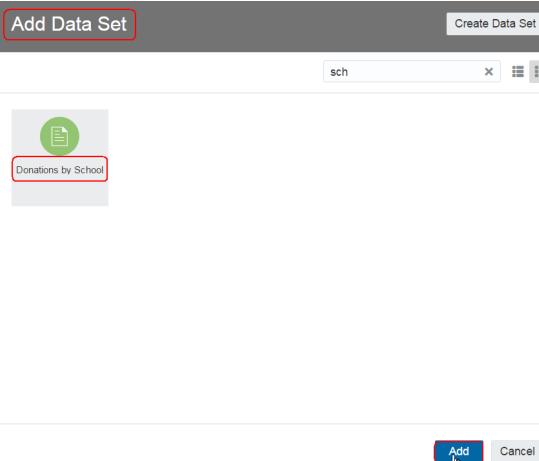
21.		<ul style="list-style-type: none"> - Select Donation by School dva project (from your local drive) - Click Import 	Select your file
22.		<ul style="list-style-type: none"> - Click on Navigator - Select Data 	Let's go to the Data region
23.		<ul style="list-style-type: none"> - 	Within the project the dependence objects as Data Sets / Data Flow(s) are automatically packaged and imported.

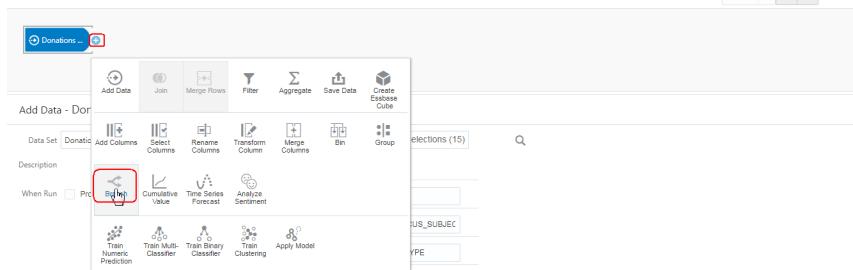
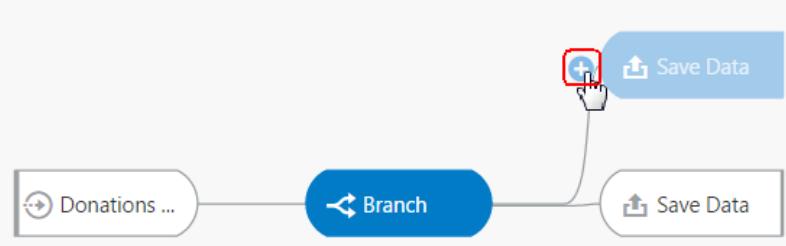
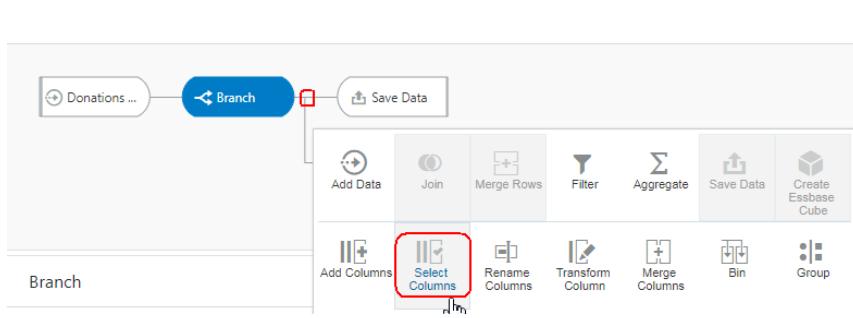
24.		<ul style="list-style-type: none"> - Go to Action Menu  - Click New Schedule 	<p>Select DF – Donations by School</p>
25.		<ul style="list-style-type: none"> - 	<p>There are various options to Schedule the data flow, according to your business needs (including Hourly)</p>
26.		<ul style="list-style-type: none"> - 	<p>Inspect it's another feature of Analytics Cloud</p>
27.		<ul style="list-style-type: none"> - Go to Action Menu  - Click New Schedule 	

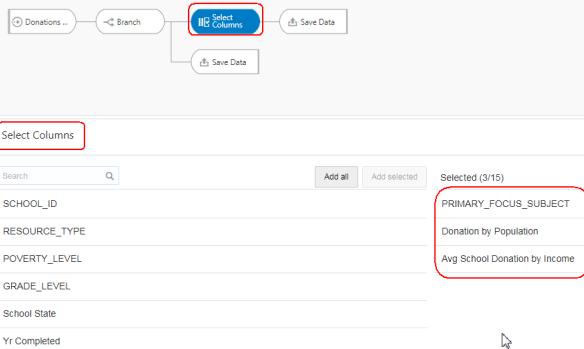
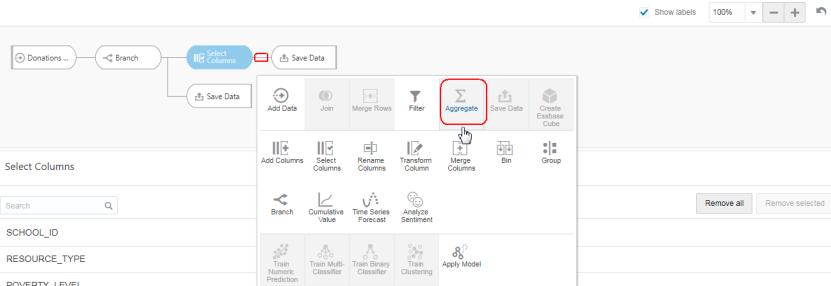
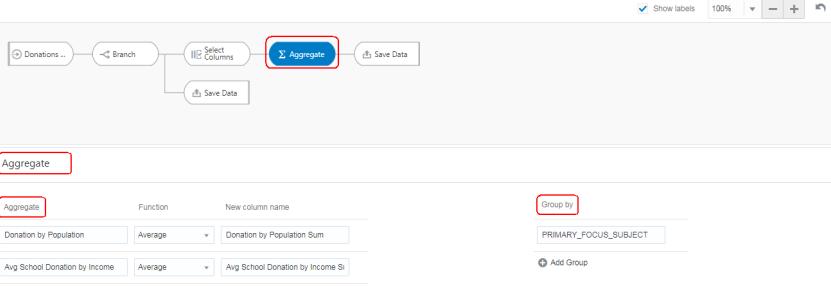
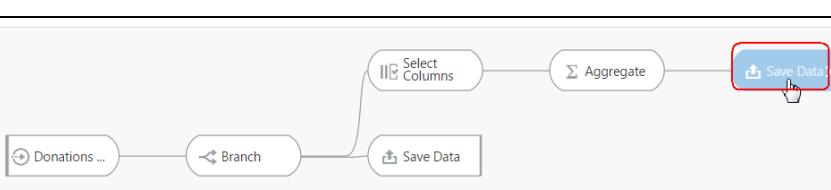
28.		-	General tab contains summary information about the Data Flow
29.		-	Inspect tab contains information related to the source data sets and the target ones
30.		-	Schedules contain information related to the Data Flow scheduler (no information means there are no schedules so far)
31.		-	History tab contains history information
32.		-	If you click on blue labeled History task, you get Jobs details

33.			General job information
34.			Information about Job Status
35.			Information about Job Parameters if any
36.			If you want to go back to History , click on < .

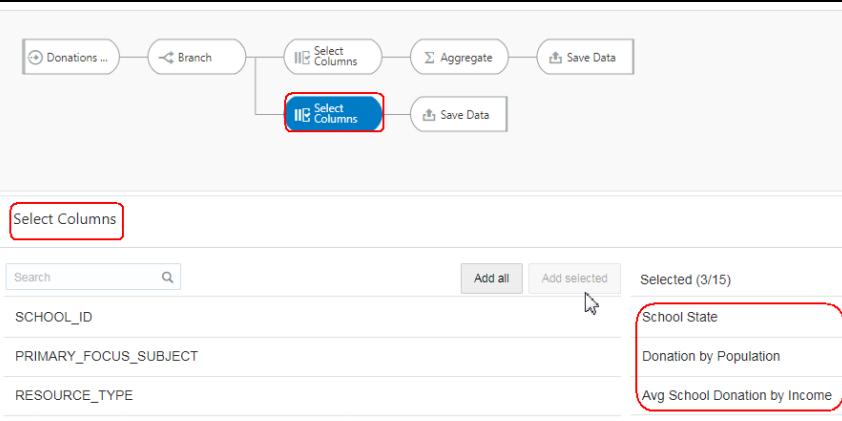
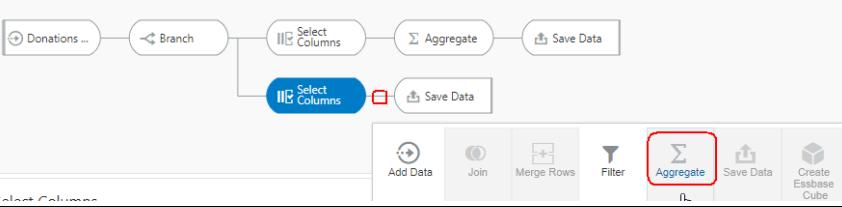
5. Create Multiple Target Data Set Using a Branch in a Data Flow

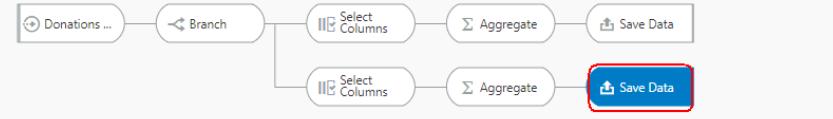
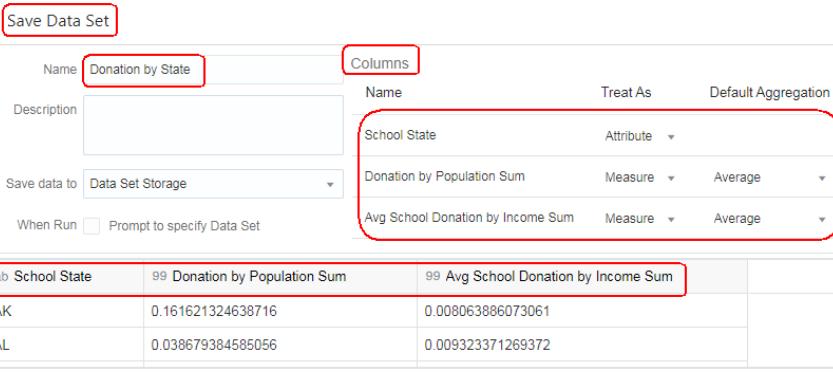
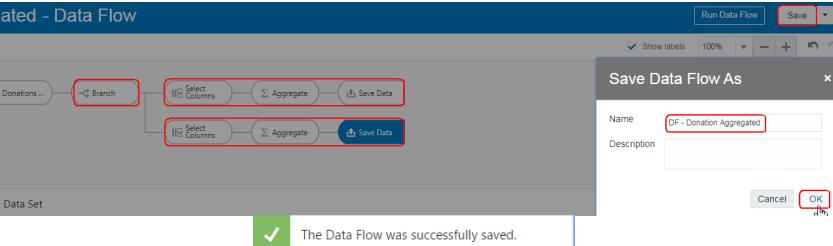
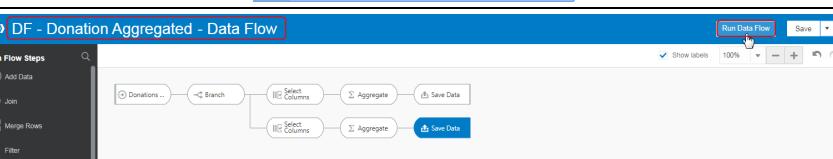
Step	View	Click stream	Talk stream
1.			<p>Let's now create a second dataflow that starts with the output of the first dataflow and generate two aggregated datasets out of it</p> <p>Once we have these two dataflows, let's see how we can execute them in a sequence</p>
1.		Click Create > Data Flow	
2.			<p>Add Data Set Donations by School</p>

Step	View	Click stream	Talk stream
2.		<ul style="list-style-type: none"> - Click on  - Click on Branch  	<p>Add Branch A branch step allows a dataflow to breakout into multiple target datasets.</p> <p>Up to 5 branches are possible in a single step.</p>
3.		<ul style="list-style-type: none"> - Click on  	<p>Let's now add a Select Columns step to choose only a few columns.</p>
4.		<ul style="list-style-type: none"> - Click on Select Columns 	

Step	View	Click stream	Talk stream
5.		<ul style="list-style-type: none"> - Click on Remove All - Select: <ul style="list-style-type: none"> o PRIMARY_FOCUS_SUBJECT o Donation by Population o Avg School Donation by Income - Click Add selected 	
6.		<ul style="list-style-type: none"> - Click on + - Select Aggregate 	Next, let's aggregate this branch
7.			Leave the defaults
8.		<ul style="list-style-type: none"> - Click on Save Data 	Save Data

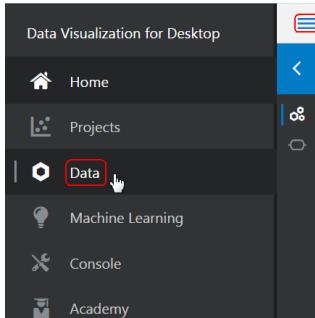
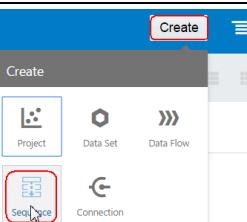
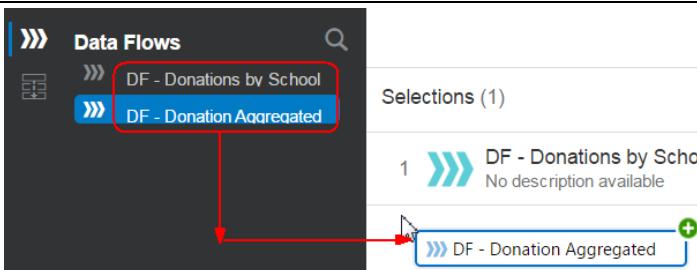
Step	View	Click stream	Talk stream																												
9.	 Save Data Set <table border="1"> <tr> <td>Name: Donation by Primary Subject Area</td> <td>Columns</td> <td>Treat As</td> <td>Default Aggregation</td> </tr> <tr> <td>Description:</td> <td>PRIMARY_FOCUS SUBJECT</td> <td>Attribute</td> <td></td> </tr> <tr> <td>Save data to: Data Set Storage</td> <td>Donation by Population Sum</td> <td>Measure</td> <td>Average</td> </tr> <tr> <td>When Run: Prompt to specify Data Set</td> <td>Avg School Donation by Income Sum</td> <td>Measure</td> <td>Average</td> </tr> </table> <table border="1"> <thead> <tr> <th>ab PRIMARY_FOCUS_SUBJECT</th> <th>99 Donation by Population Sum</th> <th>99 Avg School Donation by Income Sum</th> </tr> </thead> <tbody> <tr> <td>Applied Sciences</td> <td>0.064519920964586</td> <td>0.010433853525732</td> </tr> <tr> <td>Character Education</td> <td>0.042956671136606</td> <td>0.010246774669332</td> </tr> <tr> <td>Civics & Government</td> <td>0.067945406069012</td> <td>0.012075418619106</td> </tr> </tbody> </table>	Name: Donation by Primary Subject Area	Columns	Treat As	Default Aggregation	Description:	PRIMARY_FOCUS SUBJECT	Attribute		Save data to: Data Set Storage	Donation by Population Sum	Measure	Average	When Run: Prompt to specify Data Set	Avg School Donation by Income Sum	Measure	Average	ab PRIMARY_FOCUS_SUBJECT	99 Donation by Population Sum	99 Avg School Donation by Income Sum	Applied Sciences	0.064519920964586	0.010433853525732	Character Education	0.042956671136606	0.010246774669332	Civics & Government	0.067945406069012	0.012075418619106	<ul style="list-style-type: none"> - Type Name: Donation by Primary Subject Area 	<p>Let's name Donation by Primary Subject Area the output of this first branch.</p>
Name: Donation by Primary Subject Area	Columns	Treat As	Default Aggregation																												
Description:	PRIMARY_FOCUS SUBJECT	Attribute																													
Save data to: Data Set Storage	Donation by Population Sum	Measure	Average																												
When Run: Prompt to specify Data Set	Avg School Donation by Income Sum	Measure	Average																												
ab PRIMARY_FOCUS_SUBJECT	99 Donation by Population Sum	99 Avg School Donation by Income Sum																													
Applied Sciences	0.064519920964586	0.010433853525732																													
Character Education	0.042956671136606	0.010246774669332																													
Civics & Government	0.067945406069012	0.012075418619106																													
10.			<p>Let's select columns and aggregate in the second branch as well</p>																												
11.	 Branch * Branch into: 2 <table border="1"> <tr> <td>Add Data</td> <td>Join</td> <td>Merge Rows</td> <td>Filter</td> <td>Aggregate</td> <td>Save Data</td> <td>Create Essbase Cube</td> </tr> <tr> <td>Add Columns</td> <td>Select Column</td> <td>Rename Columns</td> <td>Transform Column</td> <td>Merge Columns</td> <td>Bin</td> <td>Group</td> </tr> <tr> <td>Branch</td> <td>Cumulative Value</td> <td>Time Series Forecast</td> <td>Analyze Sentiment</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Train Numeric Prediction</td> <td>Train Multi-Classifier</td> <td>Train Binary Classifier</td> <td>Train Clustering</td> <td>Apply Model</td> <td></td> <td></td> </tr> </table>	Add Data	Join	Merge Rows	Filter	Aggregate	Save Data	Create Essbase Cube	Add Columns	Select Column	Rename Columns	Transform Column	Merge Columns	Bin	Group	Branch	Cumulative Value	Time Series Forecast	Analyze Sentiment				Train Numeric Prediction	Train Multi-Classifier	Train Binary Classifier	Train Clustering	Apply Model			<ul style="list-style-type: none"> - Click on - Click on Select Columns 	<p>Let's now add a Select Columns step to choose only a few columns.</p>
Add Data	Join	Merge Rows	Filter	Aggregate	Save Data	Create Essbase Cube																									
Add Columns	Select Column	Rename Columns	Transform Column	Merge Columns	Bin	Group																									
Branch	Cumulative Value	Time Series Forecast	Analyze Sentiment																												
Train Numeric Prediction	Train Multi-Classifier	Train Binary Classifier	Train Clustering	Apply Model																											
12.		-																													

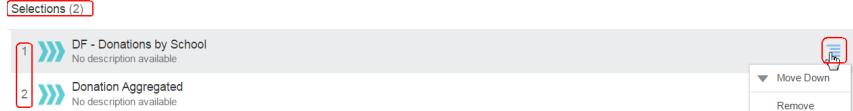
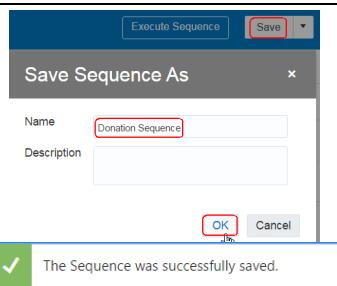
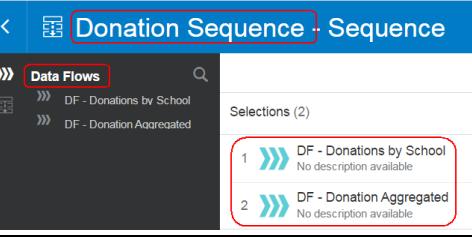
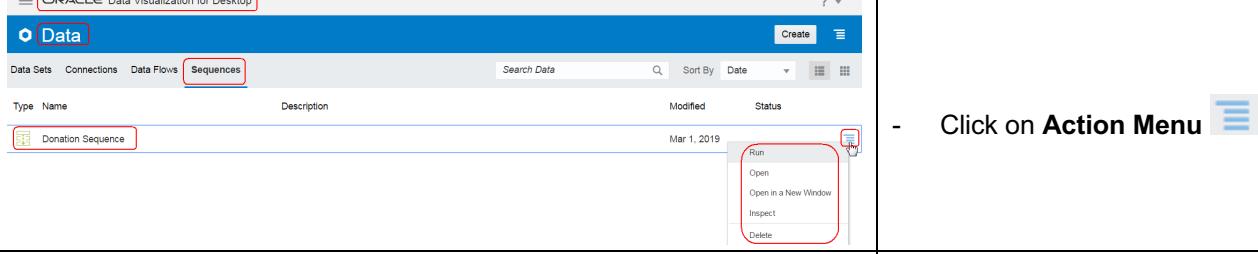
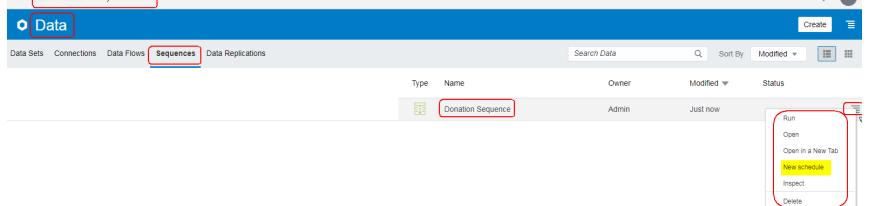
Step	View	Click stream	Talk stream
13.	 <p>Select Columns</p> <p>Search <input type="text"/> Add all Add selected Selected (3/15)</p> <ul style="list-style-type: none"> SCHOOL_ID PRIMARY_FOCUS SUBJECT RESOURCE_TYPE <p>School State Donation by Population Avg School Donation by Income</p>	<ul style="list-style-type: none"> - Click on Remove All - Select: <ul style="list-style-type: none"> ◦ School State ◦ Donation by Population ◦ Avg School Donation by Income - Click Add selected 	
14.	 <p>Select Columns</p> <p>Add Data Join Merge Rows Filter Aggregate Save Data Create Essbase Cube</p>	<ul style="list-style-type: none"> - Click on - Select Aggregate 	Next, let's aggregate this branch
15.		<ul style="list-style-type: none"> - Click on Save Data 	Save Data

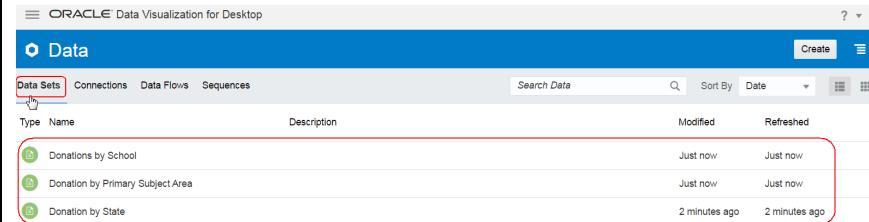
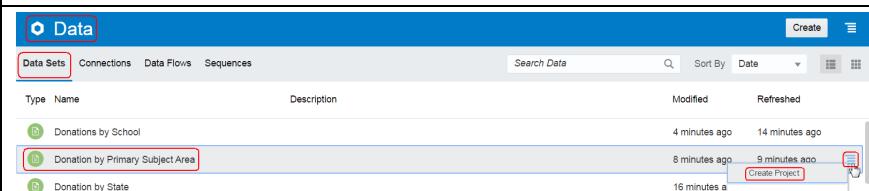
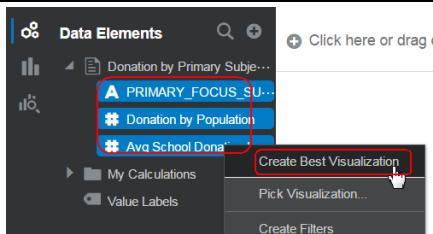
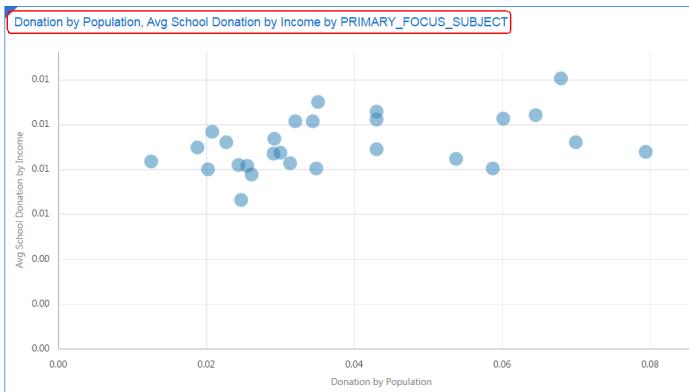
Step	View	Click stream	Talk stream
16.	 	<ul style="list-style-type: none"> Type Name: Donation by State 	<p>Let's name Donation by State the output of this second branch.</p>
17.		<ul style="list-style-type: none"> Click Save Type Name: Donation Aggregated Click Ok 	<p>Save Data Flow Donation Aggregated</p>
18.		<ul style="list-style-type: none"> Click Run Data Flow 	<p>Run Data Flow DF - Donation Aggregated</p>
19.		<ul style="list-style-type: none"> Click on  	<p>Go Back</p>
20.		<ul style="list-style-type: none"> - 	<p>You can see the two Data Flows</p>

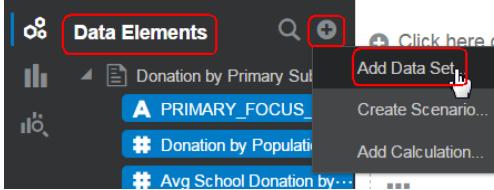
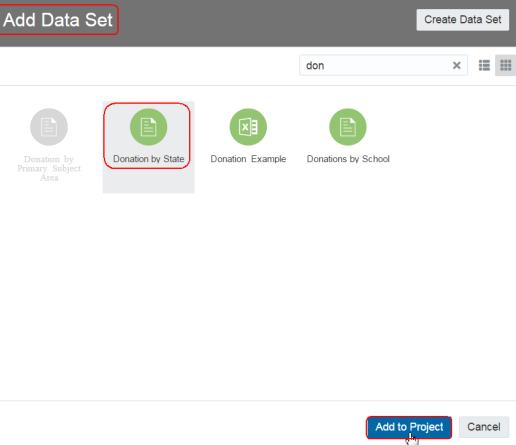
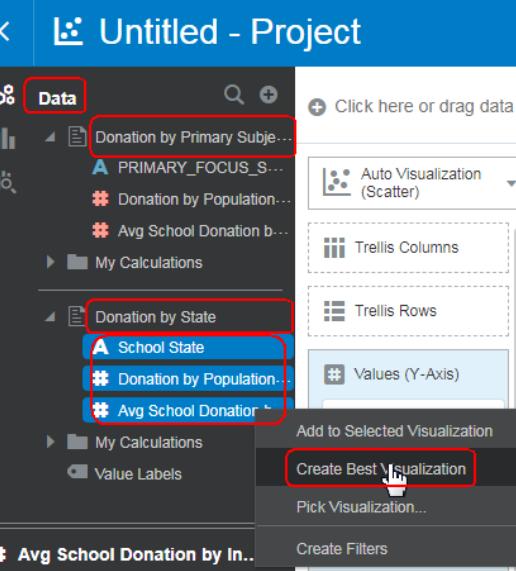
Step	View	Click stream	Talk stream														
21.		- Click on Data Sets tab	Go to Data Sets tab														
22.	 <p>The screenshot shows the Oracle Analytics interface with the 'Data' tab selected. Below the tabs, there's a header with 'Data Sets', 'Connections', 'Data Flows', 'Sequences', and 'Data Replications'. A table lists two data sets:</p> <table border="1"> <thead> <tr> <th>Type</th> <th>Name</th> <th>Owner</th> <th>Modified</th> <th>Refreshed</th> </tr> </thead> <tbody> <tr> <td>File</td> <td>Donation by State</td> <td>Admin</td> <td>5 minutes ago</td> <td>8 minutes ago</td> </tr> <tr> <td>File</td> <td>Donation by Primary Subject Area</td> <td>Admin</td> <td>Just now</td> <td>8 minutes ago</td> </tr> </tbody> </table>	Type	Name	Owner	Modified	Refreshed	File	Donation by State	Admin	5 minutes ago	8 minutes ago	File	Donation by Primary Subject Area	Admin	Just now	8 minutes ago	<p>Running DF - Donation Aggregated Data Flow will create 2 new Data Sets</p>
Type	Name	Owner	Modified	Refreshed													
File	Donation by State	Admin	5 minutes ago	8 minutes ago													
File	Donation by Primary Subject Area	Admin	Just now	8 minutes ago													

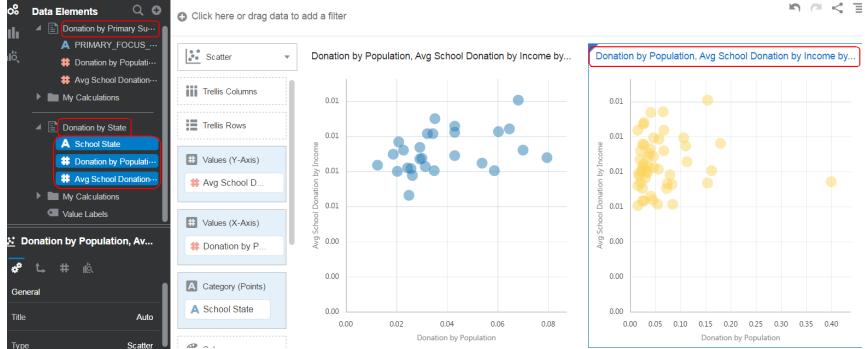
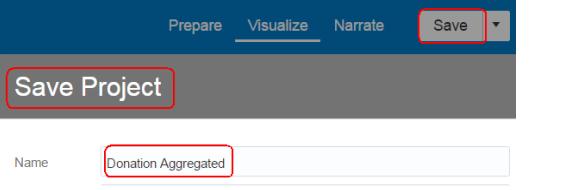
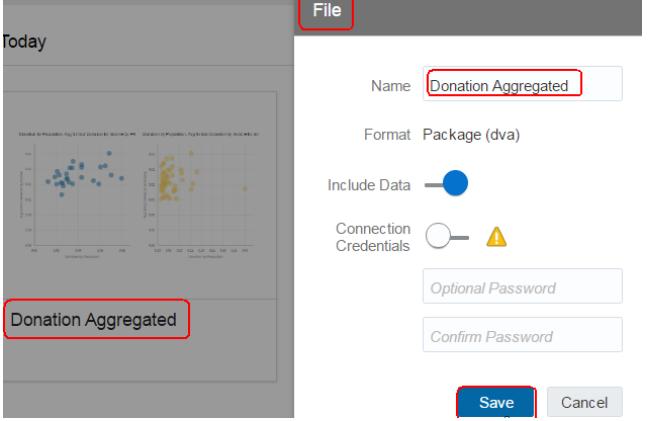
6. Advanced Properties of Data Flow

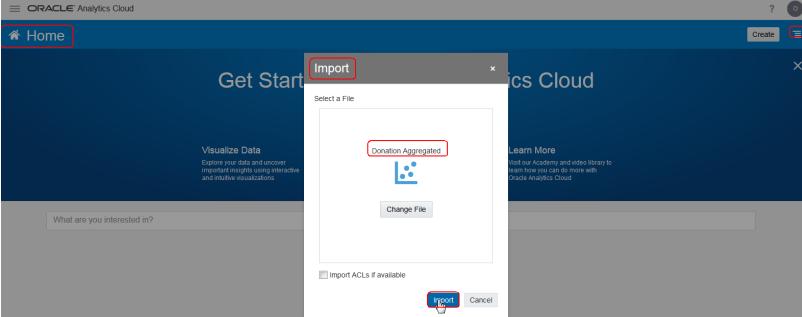
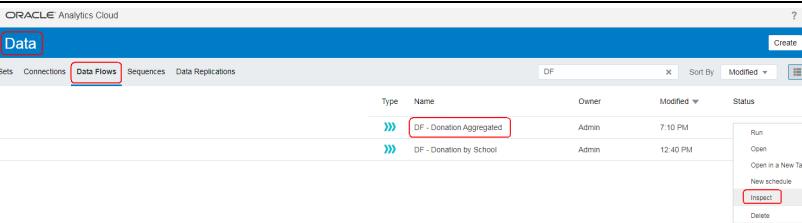
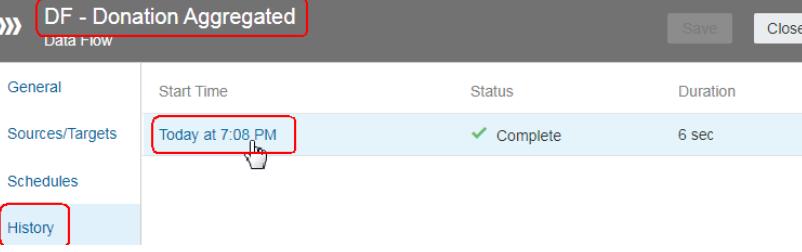
Step	View	Click stream	Talk stream
1.			<p>Let's now create a Sequence with two Data Flows.</p> <p>Sequences are containers that execute multiple dataflows one after another in a specific order.</p>
2.		<ul style="list-style-type: none"> - Click on Navigator - Click on Data 	
3.		<ul style="list-style-type: none"> - Click on Sequences 	
4.		<ul style="list-style-type: none"> - Click on Create - Select Sequence 	
5.		<ul style="list-style-type: none"> - Select DF - Donation by School and drop it to Selection area - The same process for DF - Donation Aggregates 	<p>You just drag and drop the data flows in the sequence we need them to be ran</p> <p>DF - Donation by School DF - Donation Aggregates</p>

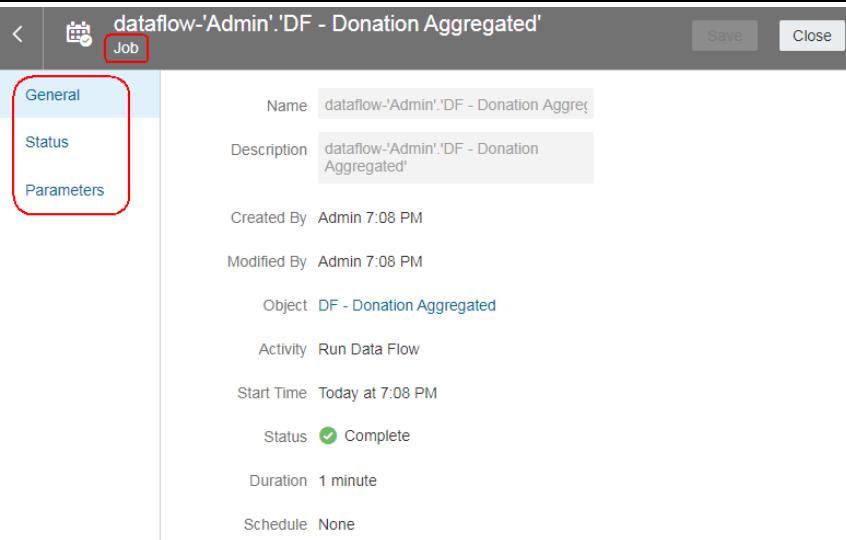
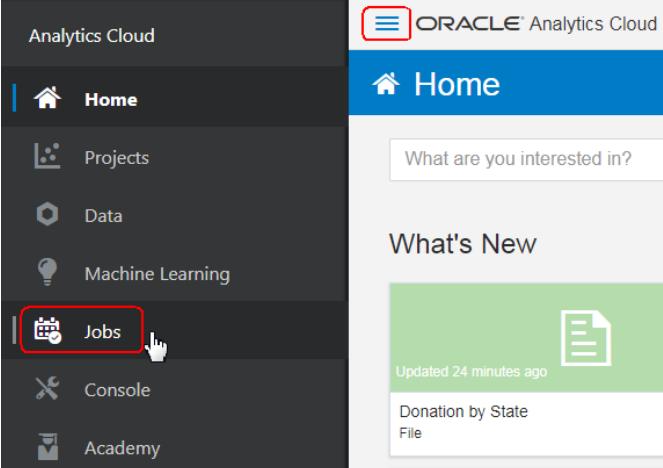
Step	View	Click stream	Talk stream
6.			If you select one DF and click on  you have the option to Move Down or Remove
7.		<ul style="list-style-type: none"> - Click on Save - Type Name: Donation Sequence - Click OK 	Save Sequence as Donation Sequence
8.		<ul style="list-style-type: none"> - Click on Execute Sequence 	Execute Sequence
9.		<ul style="list-style-type: none"> - Click on Go Back  	
10.		<ul style="list-style-type: none"> - Click on Action Menu  	You have a number of options available
11.			<p>Optional:</p> <p>With Oracle Analytics; just like dataflows, Sequences can also be scheduled for execution</p>

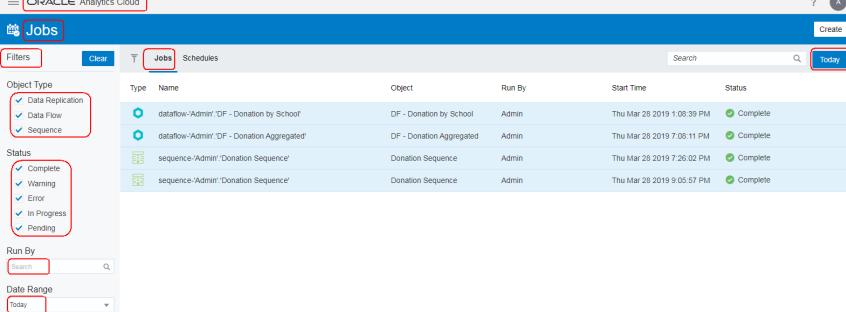
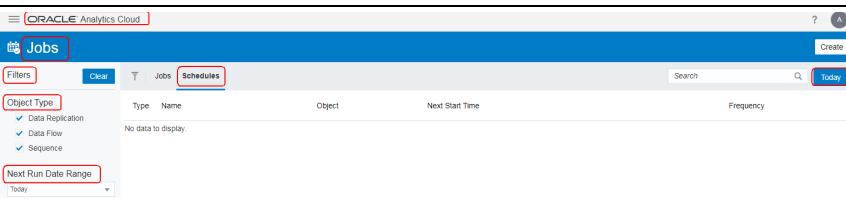
Step	View	Click stream	Talk stream
12.		- Click on Data Sets tab	These three datasets were just created by the Sequence execution which executed the two dataflows
13.		-	Let's run a quick analysis on these new datasets
14.		<ul style="list-style-type: none"> - Click on Action Menu  - Select Donation by Primary Subject Area - Select Create Project 	
15.		-	Use the print screen as guidance
16.		-	Donation by Population, Avg School Donation by PRIMARY_FOCUS SUBJECT
17.			Bring in the second aggregated dataset

Step	View	Click stream	Talk stream
18.	 <p>The screenshot shows the 'Data Elements' section of the Oracle Analytics interface. A context menu is open over the 'Add Data Set...' option, which is highlighted with a red box. Other options visible in the menu include 'Create Scenario...' and 'Add Calculation...'. The main list contains items like 'PRIMARY_FOCUS_S...', 'Donation by Population...', 'Avg School Donation by...', and 'Donation by Primary Subj...'. The 'Add Data Set' button is also highlighted with a red box.</p>		Add dataset
19.	 <p>The screenshot shows the 'Add Data Set' dialog box. The 'Create Data Set' tab is selected. In the list of datasets, 'Donation by State' is highlighted with a red box. Other options include 'Donation by Primary Subject Area', 'Donation Example', and 'Donations by School'. At the bottom right of the dialog are 'Add to Project' and 'Cancel' buttons.</p>		Donation by State
20.	 <p>The screenshot shows the 'Untitled - Project' screen. The 'Data' section is active, showing a list of data elements. 'Donation by State' is selected and highlighted with a red box. Other selected items include 'School State', 'Donation by Population...', and 'Avg School Donation...'. A context menu is open over these selected items, with 'Create Best Visualization' highlighted with a red box. Other options in the menu include 'Auto Visualization (Scatter)', 'Trellis Columns', 'Trellis Rows', and 'Values (Y-Axis)'. The main list includes 'PRIMARY_FOCUS_S...', 'Donation by Population...', 'Avg School Donation by...', and 'Donation by Primary Subj...'. At the bottom left, there's a note about 'Value Labels' and 'Create Filters'.</p>		Add a Second Visualization Donation by Population, Avg School Donation by Income by School State

Step	View	Click stream	Talk stream
21.			
22.			Save the Project as Donation Aggregated
23.			Export Donation Aggregated project

Step	View	Click stream	Talk stream
24.			and import to Oracle Analytics to highlight Job feature
25.			There are two options to go to Job Console: 1. Through Inspect Data Flow 2. Through Jobs
26.			You go to JData > Data Flows > select DF - Donation Aggregated > Inspect
27.			Choose History and click on blue labeled history items

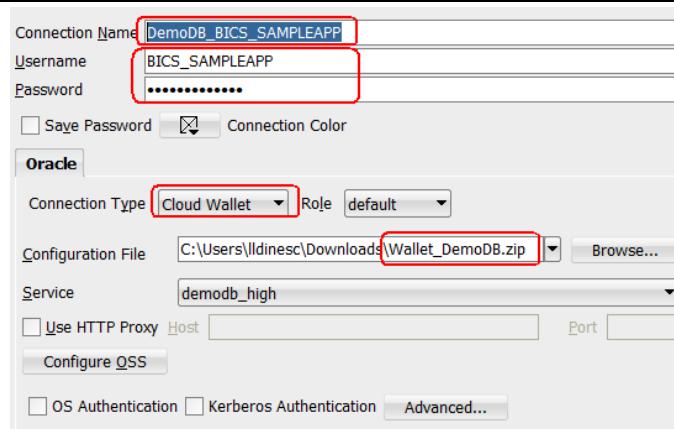
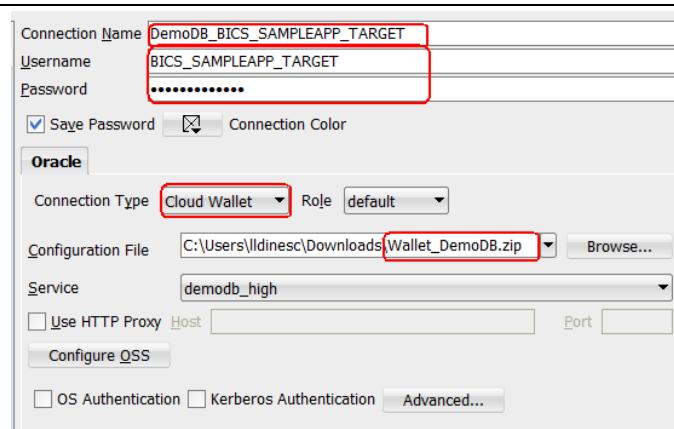
Step	View	Click stream	Talk stream
28.	 <p>Name: dataflow-'Admin'.DF - Donation Aggregated Description: dataflow-'Admin'.DF - Donation Aggregated Created By: Admin 7:08 PM Modified By: Admin 7:08 PM Object: DF - Donation Aggregated Activity: Run Data Flow Start Time: Today at 7:08 PM Status: Complete Duration: 1 minute Schedule: None</p>		You can check General , Status or Parameters tab
29.	 <p>The second option is to choose Jobs from Home Page Navigator</p>		
30.			

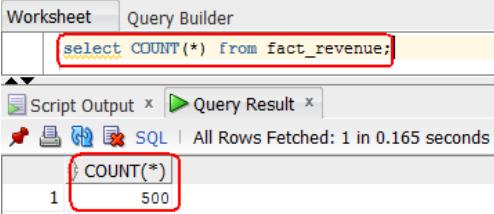
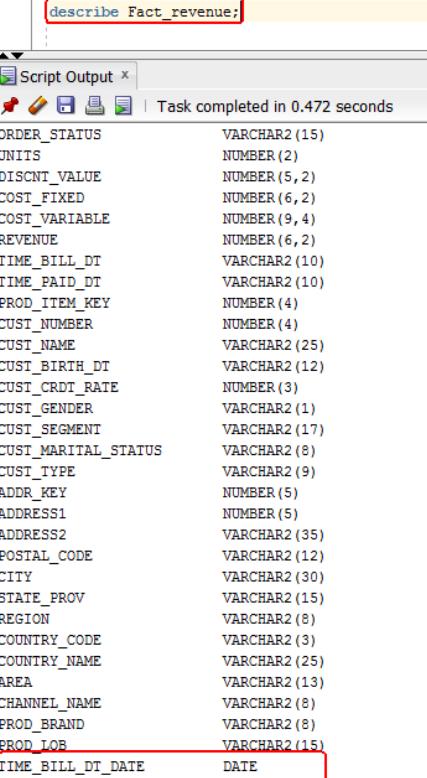
Step	View	Click stream	Talk stream
31.			<p>Jobs: Feature of Analytics Cloud: the Job Console allows to monitor execution status of all the Data Flows, Sequences and Data Replications (adhoc as well as scheduled executions)</p> <p>There are various filters available for Jobs as: Object Type, Status, Run By and Data Range.</p>
32.			<p>For Schedules available filters are just for Object Type and Next Run Data Range.</p>

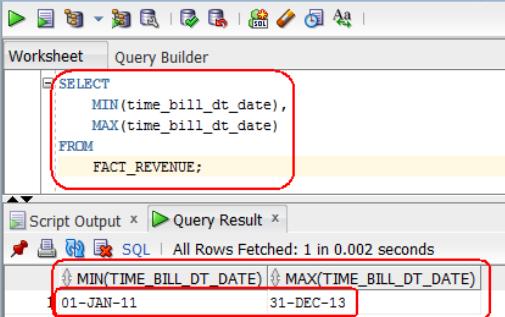
7. Working with Incremental Data with Data Flow

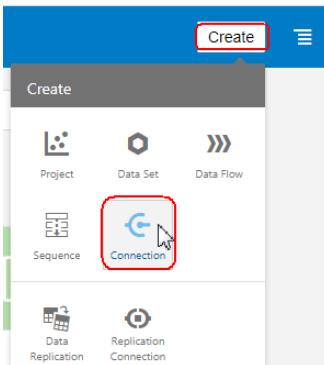
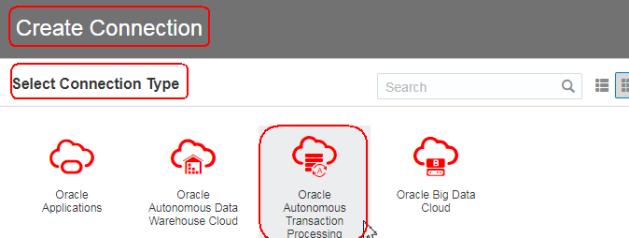
Step	View	Click stream	Talk stream
1.			<p>Incremental Data processing using Data Flows</p> <p>Data Visualization Data Flows have the ability to operate on incremental data that becomes available between batch runs.</p> <p>Incremental load greatly enhance speed (only appends new records) and also preserve history (hence a full load from a transactional data base would remove it).</p> <p>In this demo you'll see this can be configured and how one can define a method of continuing data processing from the last processed row via DV Data Flows.</p> <p>Note: This option is only available for database based Datasets Incremental Data is a feature of Analytics Cloud</p> <p>If a Data Flow consists of multiple Datasets, only one of them can be set for incremental update</p>
2.			<p>Note: The following steps are not part of the lab, but you rather can use it as a guidance, if prerequisites are met</p>
3.			<p>Prerequisites:</p> <ul style="list-style-type: none"> - You need to have access to an Analytics Cloud instance - You need to have access to a database with the tables that we are using for this demo (https://docs.oracle.com/cd/E17781_01/server.112/e18804/impexp.htm#BABHFHGH) - SQL Developer - SQL knowledge
4.			Go to your database either on-premise or cloud (DBaaS)

Step	View	Click stream	Talk stream
5.			You create 2 users (must have "CREATE TABLE" and "INSERT" privileges granted): <ul style="list-style-type: none"> - BICS_SAMPLEAPP - BICS_SAMPLEAPP_TARGET
6.		<pre>CREATE USER "BICS_SAMPLEAPP" IDENTIFIED BY "yourpassword"; GRANT "DBA" TO "BICS_SAMPLEAPP" WITH ADMIN OPTION; GRANT "CONNECT" TO "BICS_SAMPLEAPP" WITH ADMIN OPTION; GRANT "RESOURCE" TO "BICS_SAMPLEAPP" WITH ADMIN OPTION;</pre>	You grant roles
7.		<pre>CREATE USER " BICS_SAMPLEAPP_TARGET" IDENTIFIED BY "yourpassword"; GRANT "DBA" TO "BICS_SAMPLEAPP_TARGET" WITH ADMIN OPTION; GRANT "CONNECT" TO "BICS_SAMPLEAPP_TARGET" WITH ADMIN OPTION; GRANT "RESOURCE" TO "BICS_SAMPLEAPP_TARGET" WITH ADMIN OPTION;</pre>	You grant roles

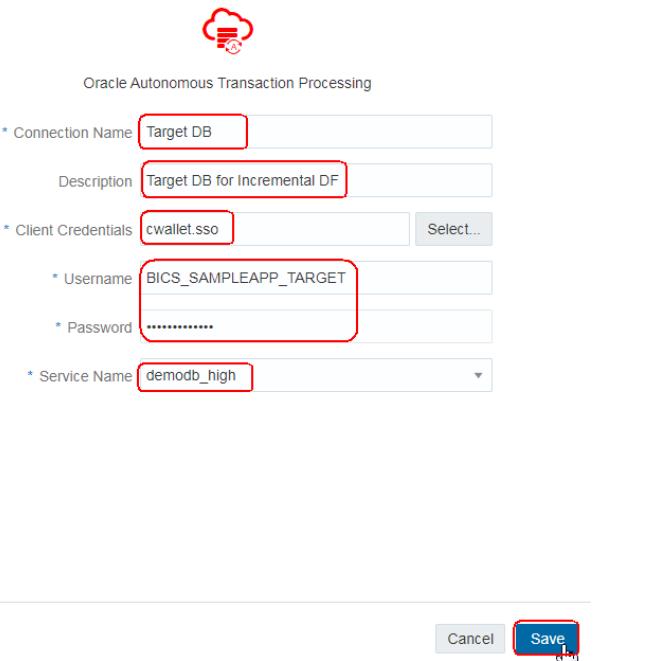
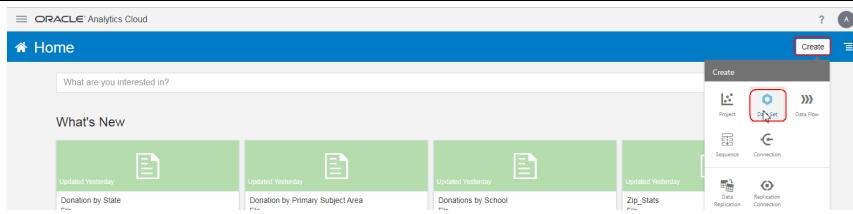
Step	View	Click stream	Talk stream
8.	 <p>Connection Name: DemoDB_BICS_SAMPLEAPP Username: BICS_SAMPLEAPP Password: <input type="checkbox"/> Save Password <input checked="" type="checkbox"/> Connection Color</p> <p>Oracle</p> <p>Connection Type: Cloud Wallet Role: default Configuration File: C:\Users\ldinesc\Downloads\Wallet_DemoDB.zip <input type="button" value="Browse..."/> Service: demodb_high <input type="checkbox"/> Use HTTP Proxy Host: Port: <input type="button" value="Configure OSS"/> <input type="checkbox"/> OS Authentication <input type="checkbox"/> Kerberos Authentication <input type="button" value="Advanced..."/></p>	-	Define two new (SQL Developer) connections for users
9.	 <p>Connection Name: DemoDB_BICS_SAMPLEAPP_TARGET Username: BICS_SAMPLEAPP_TARGET Password: <input checked="" type="checkbox"/> Save Password <input checked="" type="checkbox"/> Connection Color</p> <p>Oracle</p> <p>Connection Type: Cloud Wallet Role: default Configuration File: C:\Users\ldinesc\Downloads\Wallet_DemoDB.zip <input type="button" value="Browse..."/> Service: demodb_high <input type="checkbox"/> Use HTTP Proxy Host: Port: <input type="button" value="Configure OSS"/> <input type="checkbox"/> OS Authentication <input type="checkbox"/> Kerberos Authentication <input type="button" value="Advanced..."/></p>	-	
10.		-	Connect to DemoDB_BICS_SAMPLEAPP
11.		-	Create FACT_REVENUE table and insert records Note: copy & paste scripts from CREATE_TABLE_FACT_REVENUE.txt INSERT_INTO_FACT_REVENUE.txt If the table exist you should truncate it

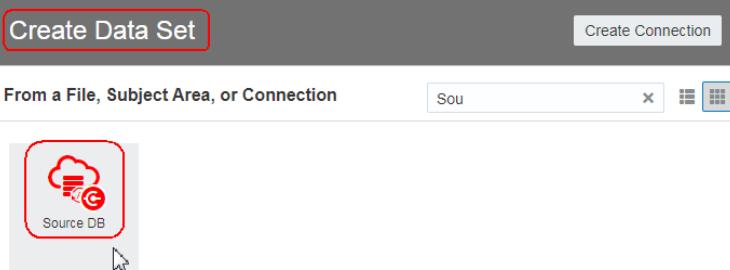
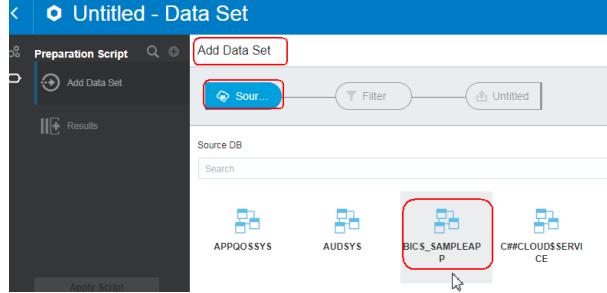
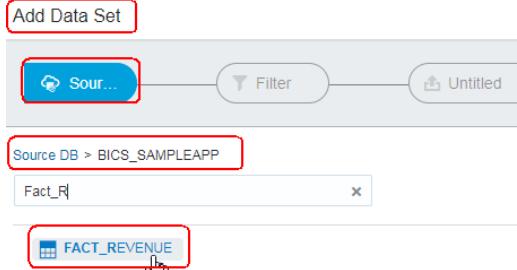
Step	View	Click stream	Talk stream																																																																
12.	 <pre>select COUNT(*) from fact_revenue;</pre> <p>Script Output x Query Result x SQL All Rows Fetched: 1 in 0.165 seconds</p> <table border="1"> <thead> <tr> <th>COUNT(*)</th> </tr> </thead> <tbody> <tr> <td>1 500</td> </tr> </tbody> </table>	COUNT(*)	1 500	select COUNT(*) from fact_revenue;	You should have inserted 500 records																																																														
COUNT(*)																																																																			
1 500																																																																			
13.	 <pre>describe Fact_revenue</pre> <p>Script Output x Task completed in 0.472 seconds</p> <table border="1"> <thead> <tr> <th>Column Name</th> <th>Data Type</th> </tr> </thead> <tbody> <tr><td>ORDER_STATUS</td><td>VARCHAR2(15)</td></tr> <tr><td>UNITS</td><td>NUMBER(2)</td></tr> <tr><td>DISCNT_VALUE</td><td>NUMBER(5,2)</td></tr> <tr><td>COST_FIXED</td><td>NUMBER(6,2)</td></tr> <tr><td>COST_VARIABLE</td><td>NUMBER(9,4)</td></tr> <tr><td>REVENUE</td><td>NUMBER(6,2)</td></tr> <tr><td>TIME_BILL_DT</td><td>VARCHAR2(10)</td></tr> <tr><td>TIME_PAID_DT</td><td>VARCHAR2(10)</td></tr> <tr><td>PROD_ITEM_KEY</td><td>NUMBER(4)</td></tr> <tr><td>CUST_NUMBER</td><td>NUMBER(4)</td></tr> <tr><td>CUST_NAME</td><td>VARCHAR2(25)</td></tr> <tr><td>CUST_BIRTH_DT</td><td>VARCHAR2(12)</td></tr> <tr><td>CUST_CRDT_RATE</td><td>NUMBER(3)</td></tr> <tr><td>CUST_GENDER</td><td>VARCHAR2(1)</td></tr> <tr><td>CUST_SEGMENT</td><td>VARCHAR2(17)</td></tr> <tr><td>CUST_MARITAL_STATUS</td><td>VARCHAR2(8)</td></tr> <tr><td>CUST_TYPE</td><td>VARCHAR2(9)</td></tr> <tr><td>ADDR_KEY</td><td>NUMBER(5)</td></tr> <tr><td>ADDRESS1</td><td>NUMBER(5)</td></tr> <tr><td>ADDRESS2</td><td>VARCHAR2(35)</td></tr> <tr><td>POSTAL_CODE</td><td>VARCHAR2(12)</td></tr> <tr><td>CITY</td><td>VARCHAR2(30)</td></tr> <tr><td>STATE_PROV</td><td>VARCHAR2(15)</td></tr> <tr><td>REGION</td><td>VARCHAR2(8)</td></tr> <tr><td>COUNTRY_CODE</td><td>VARCHAR2(3)</td></tr> <tr><td>COUNTRY_NAME</td><td>VARCHAR2(25)</td></tr> <tr><td>AREA</td><td>VARCHAR2(13)</td></tr> <tr><td>CHANNEL_NAME</td><td>VARCHAR2(6)</td></tr> <tr><td>PROD_BRAND</td><td>VARCHAR2(8)</td></tr> <tr><td>PROD_LOB</td><td>VARCHAR2(15)</td></tr> <tr><td>TIME_BILL_DT_DATE</td><td>DATE</td></tr> </tbody> </table>	Column Name	Data Type	ORDER_STATUS	VARCHAR2(15)	UNITS	NUMBER(2)	DISCNT_VALUE	NUMBER(5,2)	COST_FIXED	NUMBER(6,2)	COST_VARIABLE	NUMBER(9,4)	REVENUE	NUMBER(6,2)	TIME_BILL_DT	VARCHAR2(10)	TIME_PAID_DT	VARCHAR2(10)	PROD_ITEM_KEY	NUMBER(4)	CUST_NUMBER	NUMBER(4)	CUST_NAME	VARCHAR2(25)	CUST_BIRTH_DT	VARCHAR2(12)	CUST_CRDT_RATE	NUMBER(3)	CUST_GENDER	VARCHAR2(1)	CUST_SEGMENT	VARCHAR2(17)	CUST_MARITAL_STATUS	VARCHAR2(8)	CUST_TYPE	VARCHAR2(9)	ADDR_KEY	NUMBER(5)	ADDRESS1	NUMBER(5)	ADDRESS2	VARCHAR2(35)	POSTAL_CODE	VARCHAR2(12)	CITY	VARCHAR2(30)	STATE_PROV	VARCHAR2(15)	REGION	VARCHAR2(8)	COUNTRY_CODE	VARCHAR2(3)	COUNTRY_NAME	VARCHAR2(25)	AREA	VARCHAR2(13)	CHANNEL_NAME	VARCHAR2(6)	PROD_BRAND	VARCHAR2(8)	PROD_LOB	VARCHAR2(15)	TIME_BILL_DT_DATE	DATE		Our source table is FACT_REVENUE .
Column Name	Data Type																																																																		
ORDER_STATUS	VARCHAR2(15)																																																																		
UNITS	NUMBER(2)																																																																		
DISCNT_VALUE	NUMBER(5,2)																																																																		
COST_FIXED	NUMBER(6,2)																																																																		
COST_VARIABLE	NUMBER(9,4)																																																																		
REVENUE	NUMBER(6,2)																																																																		
TIME_BILL_DT	VARCHAR2(10)																																																																		
TIME_PAID_DT	VARCHAR2(10)																																																																		
PROD_ITEM_KEY	NUMBER(4)																																																																		
CUST_NUMBER	NUMBER(4)																																																																		
CUST_NAME	VARCHAR2(25)																																																																		
CUST_BIRTH_DT	VARCHAR2(12)																																																																		
CUST_CRDT_RATE	NUMBER(3)																																																																		
CUST_GENDER	VARCHAR2(1)																																																																		
CUST_SEGMENT	VARCHAR2(17)																																																																		
CUST_MARITAL_STATUS	VARCHAR2(8)																																																																		
CUST_TYPE	VARCHAR2(9)																																																																		
ADDR_KEY	NUMBER(5)																																																																		
ADDRESS1	NUMBER(5)																																																																		
ADDRESS2	VARCHAR2(35)																																																																		
POSTAL_CODE	VARCHAR2(12)																																																																		
CITY	VARCHAR2(30)																																																																		
STATE_PROV	VARCHAR2(15)																																																																		
REGION	VARCHAR2(8)																																																																		
COUNTRY_CODE	VARCHAR2(3)																																																																		
COUNTRY_NAME	VARCHAR2(25)																																																																		
AREA	VARCHAR2(13)																																																																		
CHANNEL_NAME	VARCHAR2(6)																																																																		
PROD_BRAND	VARCHAR2(8)																																																																		
PROD_LOB	VARCHAR2(15)																																																																		
TIME_BILL_DT_DATE	DATE																																																																		

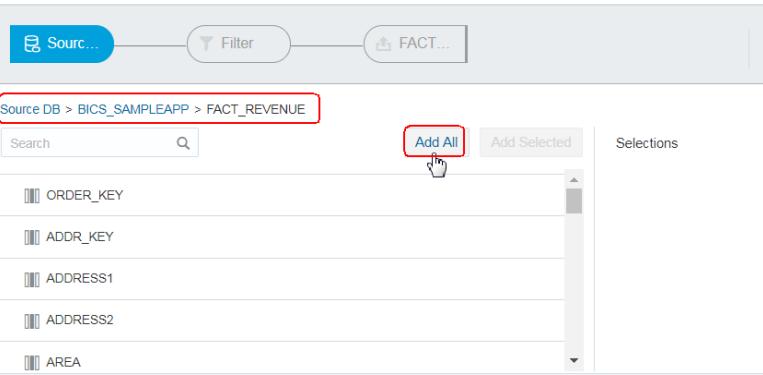
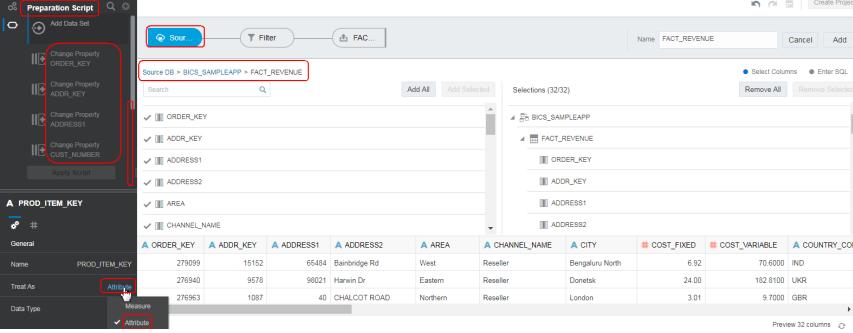
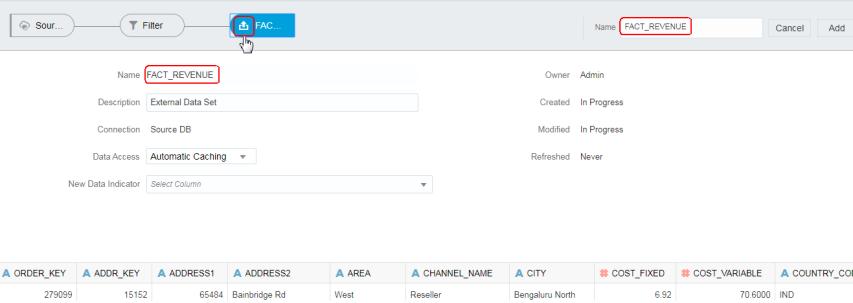
Step	View	Click stream	Talk stream				
14.	 <pre> SELECT MIN(time_bill_dt_date), MAX(time_bill_dt_date) FROM FACT_REVENUE; </pre> <p>Script Output x All Rows Fetched: 1 in 0.002 seconds</p> <table border="1"> <thead> <tr> <th>MIN(TIME_BILL_DT_DATE)</th> <th>MAX(TIME_BILL_DT_DATE)</th> </tr> </thead> <tbody> <tr> <td>01-JAN-11</td> <td>31-DEC-13</td> </tr> </tbody> </table>	MIN(TIME_BILL_DT_DATE)	MAX(TIME_BILL_DT_DATE)	01-JAN-11	31-DEC-13	<pre> SELECT MIN(time_bill_dt_date), MAX(time_bill_dt_date) FROM FACT_REVENUE; </pre>	<p>It currently has data for the years 01-JAN-2011 to 31-DEC-2013.</p> <p>TIME_BILL_DT_DATE is a column that stores the billing date for each record and will be used to identify new data qualifying for incremental loads.</p>
MIN(TIME_BILL_DT_DATE)	MAX(TIME_BILL_DT_DATE)						
01-JAN-11	31-DEC-13						
15.			Switch to Oracle Analytics instance				
16.			<p>Now you define two database connections; one to the Source DB schema, and one to an empty Target DB schema.</p> <p>To make it simple, you can use the same database (connection) with 2 users:</p> <ul style="list-style-type: none"> - First user (BICS_SAMPLEAPP) has all the source tables (FACT_REVENUE) - Second user (BICS_SAMPLEAPP_TARGET) has no tables at this point in time 				
17.			<p>Let's create Source DB connection in Oracle Analytics</p> <p>In previous section you have already created a database data source connection, so you follow the below relevant print screens</p>				

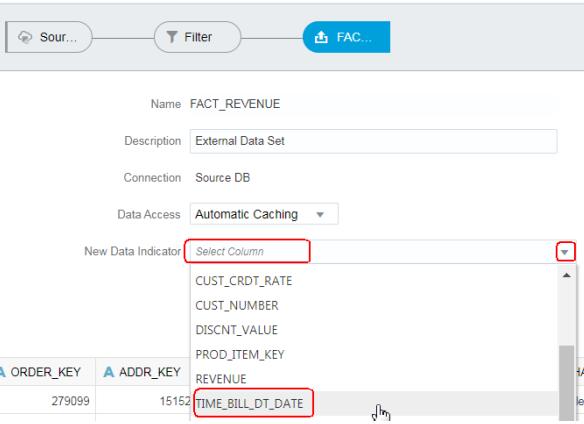
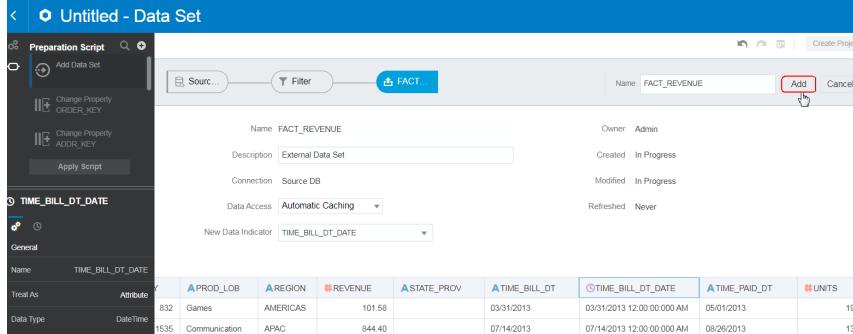
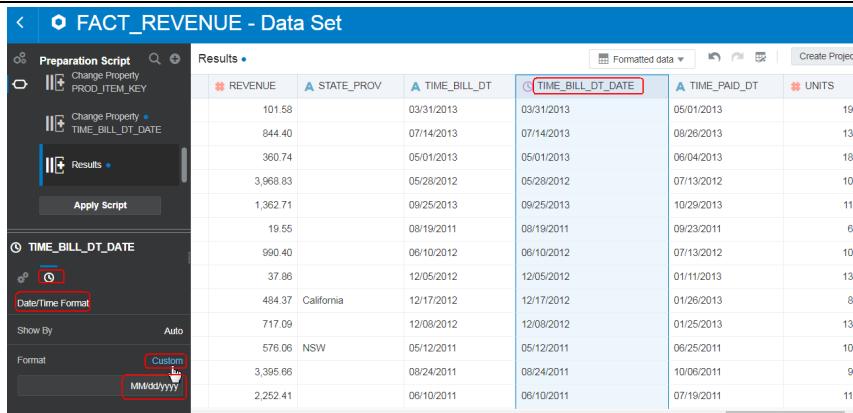
Step	View	Click stream	Talk stream
18.	 <p>The screenshot shows the Oracle Analytics interface with the 'Create' button at the top. Below it is a dropdown menu with several options: 'Project', 'Data Set', 'Data Flow', 'Sequence', 'Connection' (which is highlighted with a red box and a cursor icon), 'Data Replication', and 'Replication Connection'.</p>	<ul style="list-style-type: none"> - Click Create > Connection 	
19.	 <p>The screenshot shows the 'Create Connection' dialog. At the top is a header bar with the title 'Create Connection'. Below it is a search bar labeled 'Search' with a magnifying glass icon. Underneath the search bar is a section titled 'Select Connection Type' with a dropdown arrow icon. There are four connection types listed: 'Oracle Applications', 'Oracle Autonomous Data Warehouse Cloud', 'Oracle Autonomous Transaction Processing' (which is highlighted with a red box and a cursor icon), and 'Oracle Big Data Cloud'.</p>	<ul style="list-style-type: none"> - Select Connection Type > Oracle Autonomous Transaction Processing 	

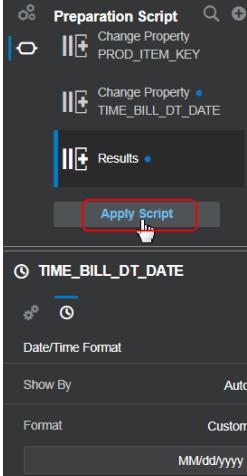
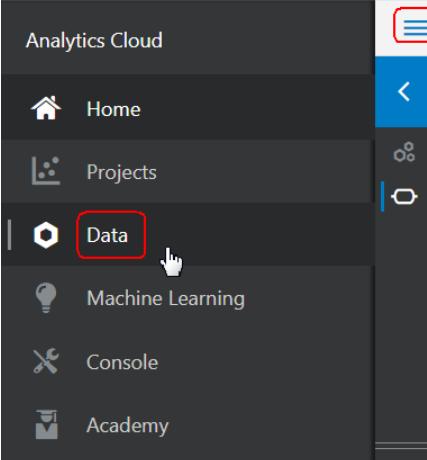
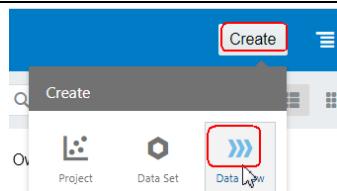
Step	View	Click stream	Talk stream
20.	<p>Create Connection</p> <p>Oracle Autonomous Transaction Processing</p> <p>* Connection Name: Source DB</p> <p>Description: Source DB for Incremental DF</p> <p>* Client Credentials: cwallet.sso Select...</p> <p>* Username: BICS_SAMPLEAPP</p> <p>* Password: [REDACTED]</p> <p>* Service Name: demodb_high</p> <p>Cancel Save</p>	-	Use your own credentials
21.		-	Let's create Target DB connection

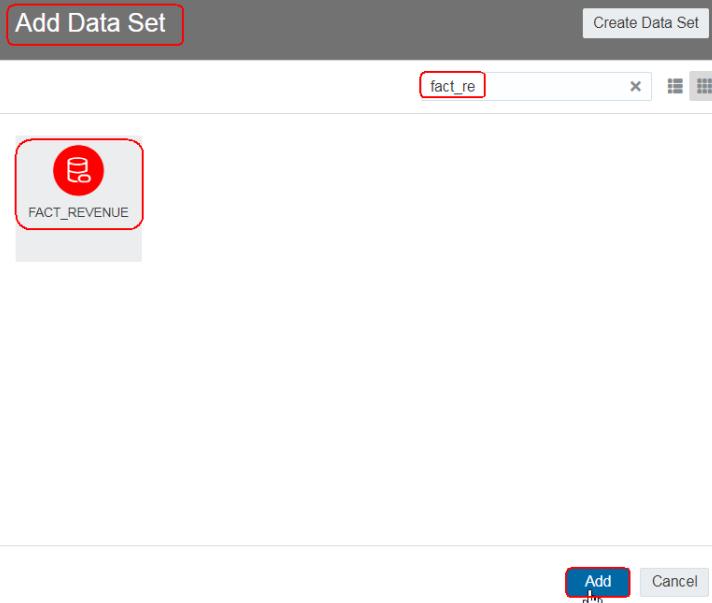
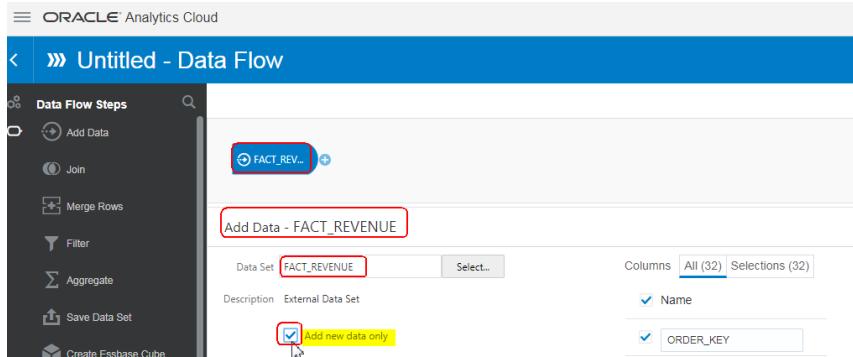
Step	View	Click stream	Talk stream
22.		-	Use your own credentials
23.		-	Let's now Create the Source Dataset FACT_REVENUE in Oracle Analytics
24.			

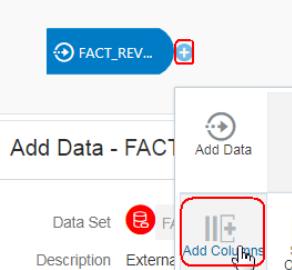
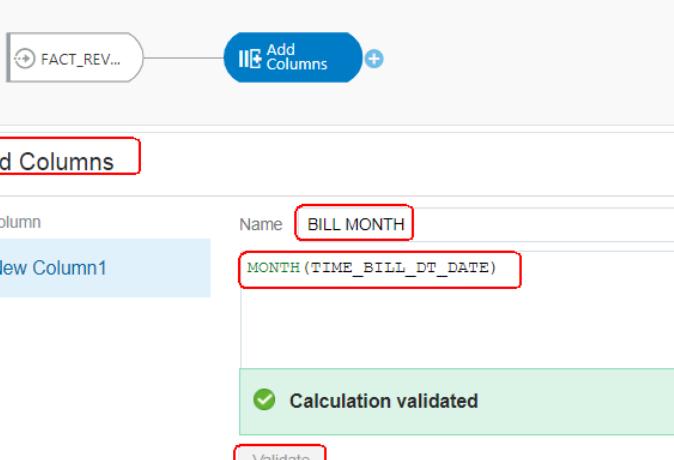
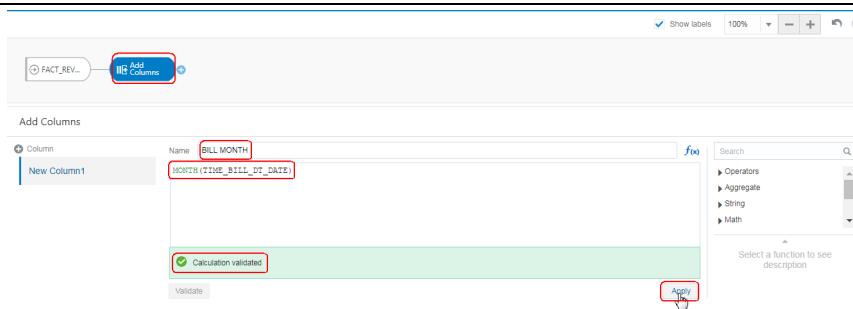
Step	View	Click stream	Talk stream
25.		- Click on Source DB	Create Source Data Set
26.		- Click on BICS_SAMPLEAPP	
27.		- Click on FACT_REVENUE	Search for FACT_REVENUE table and add all columns

Step	View	Click stream	Talk stream
28.		- Click Add All	
29.		- Click on Get Preview Data	
30.		-	<p>Change from Measure to Attribute: ORDER_KEY, ADDR_KEY, ADDRESS1, CUST_NUMBER, PROD_ITEM_KEY</p> <p>Check Preparation Script Pane</p>
31.		- Click on 	

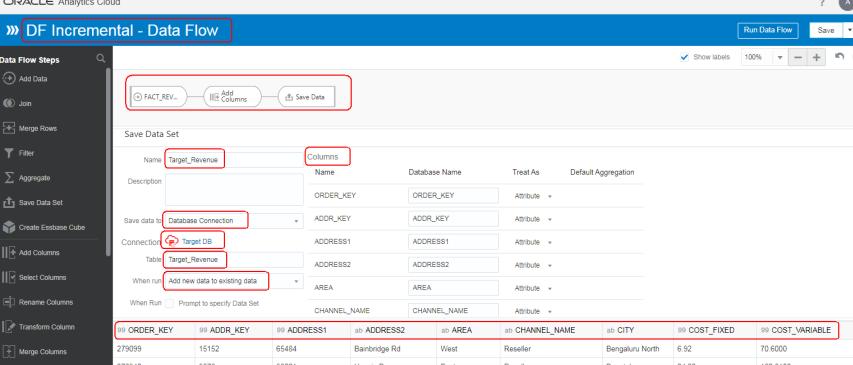
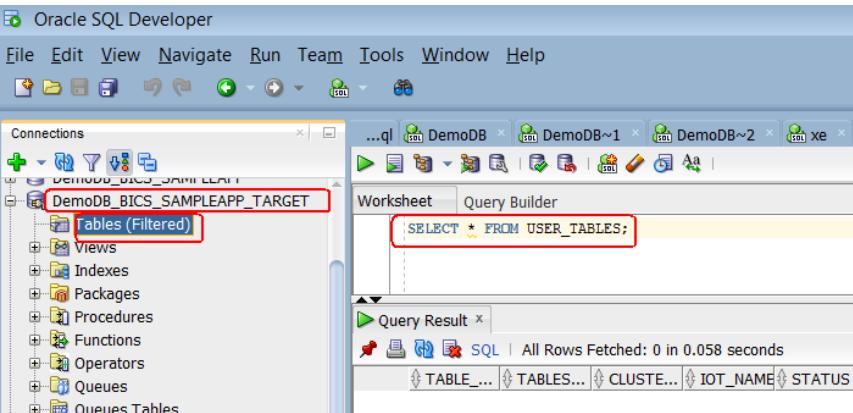
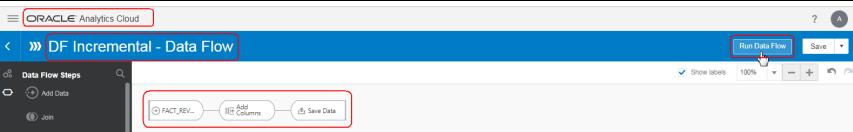
Step	View	Click stream	Talk stream
32.		<ul style="list-style-type: none"> - Click on New Data Indicator arrow ▾ and choose TIME_BILL_DT_DATE 	<p>Let's edit the data set to define a "New Data Identifier". This column acts as a cursor to identify which data was read during previous runs.</p>
33.		<ul style="list-style-type: none"> - Click Add 	<p>After you click Add, you'll get the new FACT_REVENUE Data Set</p>
34.		<ul style="list-style-type: none"> - 	<p>You can change the format for TIME_BILL_DT_DATE from "MM/dd/yyyy hh:mm:ss:SSS tt" to "MM/dd/yyyy"</p>

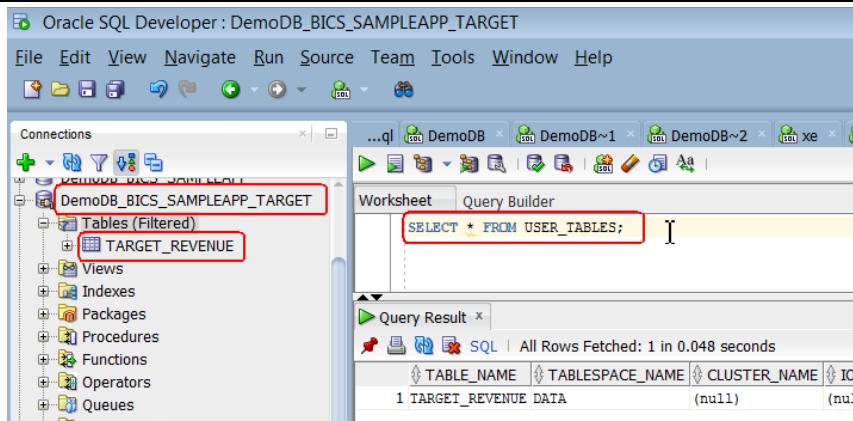
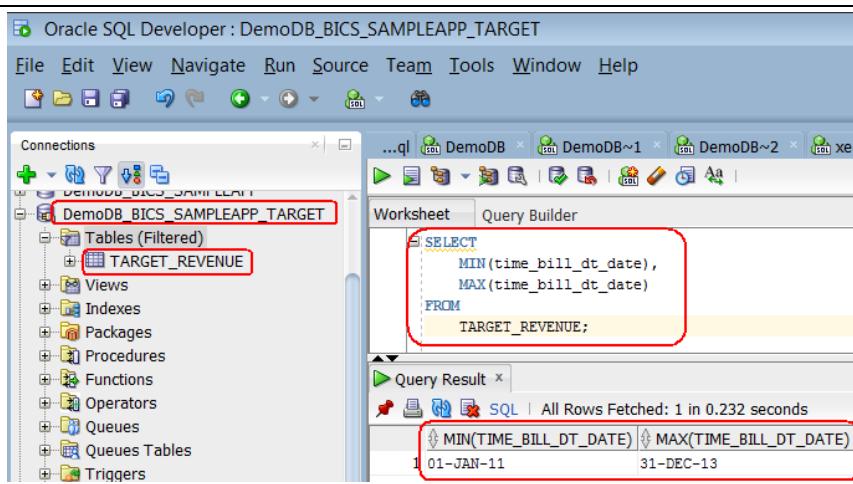
Step	View	Click stream	Talk stream
35.		- Click Apply Script	
36.		-	
37.		- Click on Navigator > Home Page	Go back to Home Page
38.		- Click on Create > Data Flow	Let's now define our (Incremental) Data Flow

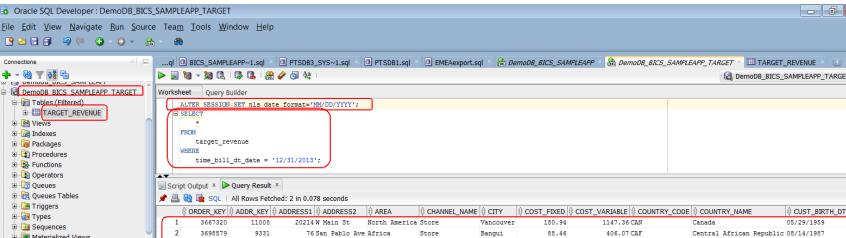
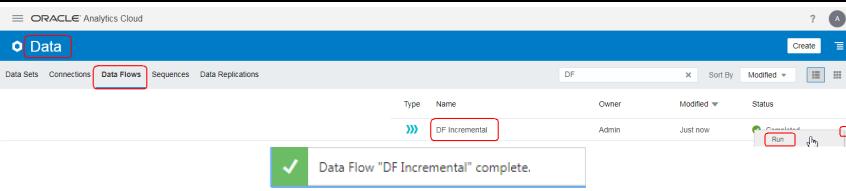
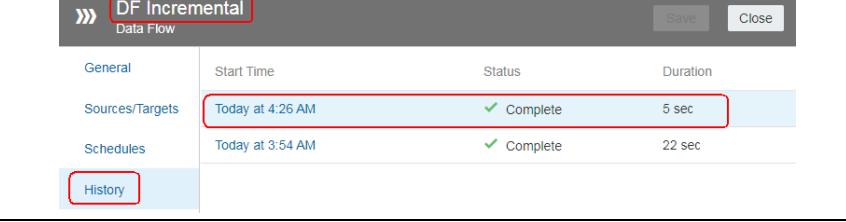
Step	View	Click stream	Talk stream
39.		<ul style="list-style-type: none"> - Search for FACT_REVENUE - Select FACT_REVENUE - Click Add 	Our source data is FACT_REVENUE Data Set
40.		<ul style="list-style-type: none"> - Check on Add new data only 	<p>Notice the flag for "Add new data only". This flag appears ONLY if a New Data Identifier was defined on the source data set.</p> <p>Checking this flag on will activate incremental processing of data in the Data Flow.</p>

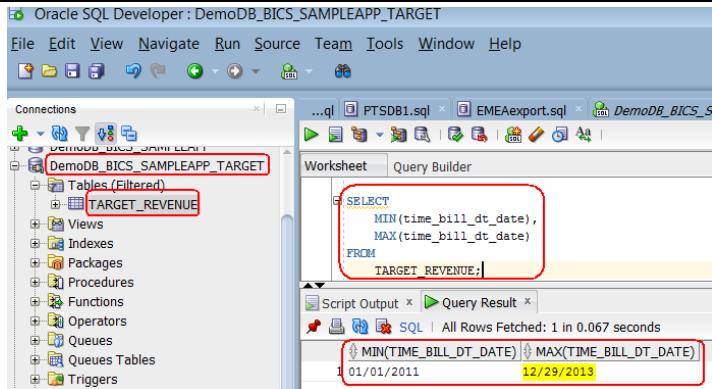
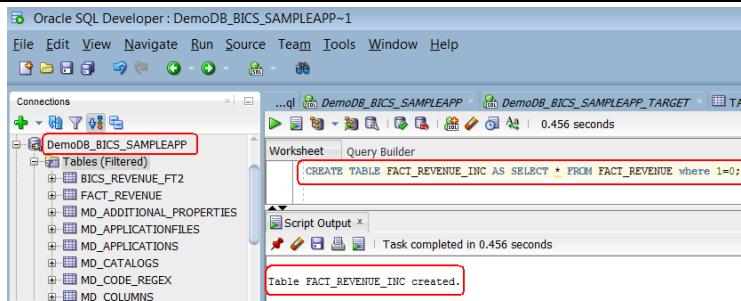
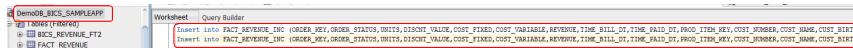
Step	View	Click stream	Talk stream
41.		<ul style="list-style-type: none"> - Click on Add a Step + - Select Add Columns 	Let's add a basic extraction node to our Data Flow, for demo purpose
42.		<ul style="list-style-type: none"> - Type Name: BILL MONTH - Type Formula: MONTH(TIME_BILL_DT_DATE) - Click on Validate 	Let's extract month
43.		<ul style="list-style-type: none"> - Click Apply 	

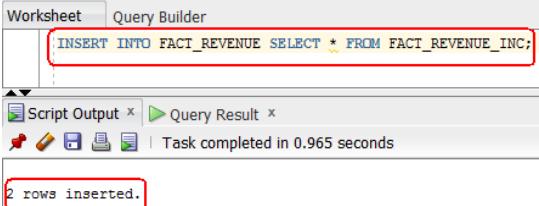
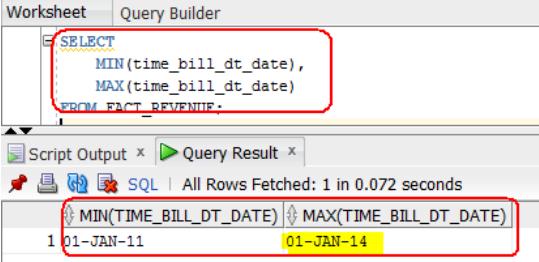
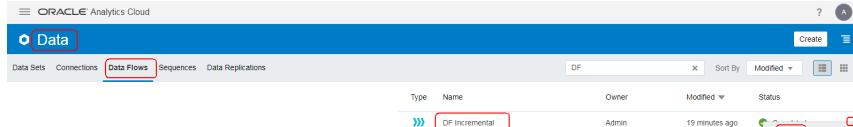
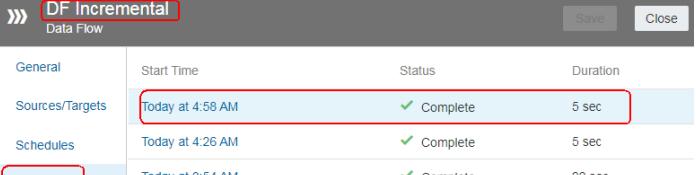
Step	View	Click stream	Talk stream
44.		<ul style="list-style-type: none"> - Click on Add a Step + - Select Save Data 	+Save Data And save the output Dataset to our target Connection
45.		<ul style="list-style-type: none"> - Type Name: Target_Revenue - Click on Save data to arrow and Select Database Connection - Click on Select Connection - Find and Select TargetDB - Type Target DB Table: Target_DB - Click on When run arrow and select Add new data to existing data 	Name: Target_Revenue Save data to: Database Connection When run: Add new data to existing data The option now appears to either replace data in target for every run, or only add new data. This is only available if target Dataset is stored to a database connection Select Connection: Target
46.		-	Save Data flow as DF Incremental
47.		<ul style="list-style-type: none"> - Type Name: DF Incremental - Click Ok 	

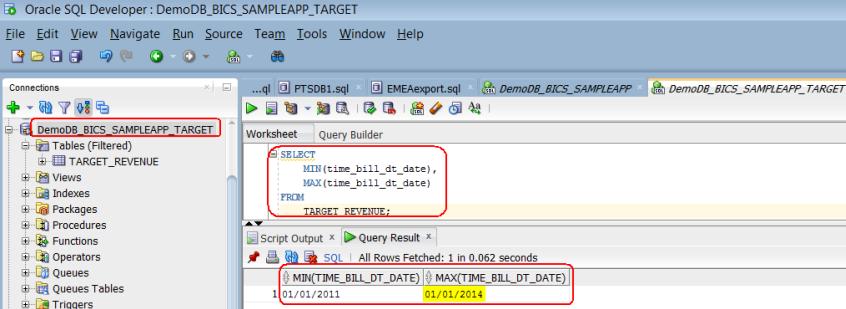
Step	View	Click stream	Talk stream
48.		-	-
49.		-	Switch to SQL Developer
50.		<pre>SELECT * FROM USER_TABLES;</pre>	<p>Let's check if we have any tables in our target database.</p> <p>There is no table at this point in time.</p>
51.		-	Switch to Analytics Cloud
52.		- Click on Run Data Flow	<p>Let's go back to our Analytics Cloud window and run our Data Flow</p>
53.		-	Switch to SQL Developer

Step	View	Click stream	Talk stream
54.	 <pre>File Edit View Navigate Run Source Team Tools Window Help Connections Worksheet Query Builder DemoDB_BICS_SAMPLEAPP_TARGET Tables (Filtered) SELECT * FROM USER_TABLES; TARGET_REVENU Views Indexes Packages Procedures Functions Operators Queues Queue Tables Tables Triggers</pre> <pre>Query Result SQL All Rows Fetched: 1 in 0.048 seconds TABLE_NAME TABLESPACE_NAME CLUSTER_NAME IOT 1 TARGET_REVENU DATA (null) (null)</pre>	-	Go to SQL Developer and check for the new table
55.		-	Target table was created in the Target Database, and since this is the first run it's mirror copy of FACT_REVENU source table
56.	 <pre>File Edit View Navigate Run Source Team Tools Window Help Connections Worksheet Query Builder DemoDB_BICS_SAMPLEAPP_TARGET Tables (Filtered) SELECT MIN(time_bill_dt_date), MAX(time_bill_dt_date) FROM TARGET_REVENU; TARGET_REVENU Views Indexes Packages Procedures Functions Operators Queues Queue Tables Triggers</pre> <pre>Query Result SQL All Rows Fetched: 1 in 0.232 seconds MIN(TIME_BILL_DT_DATE) MAX(TIME_BILL_DT_DATE) 1 01-JAN-11 31-DEC-13</pre>	<pre>SELECT MIN(time_bill_dt_date), MAX(time_bill_dt_date) FROM TARGET_REVENU;</pre>	It currently has data for the years 01-JAN-2011 to 31-DEC-2013.

Step	View	Click stream	Talk stream
57.		<pre>ALTER SESSION SET nls_date_format='MM/DD/YYYY'; SELECT * FROM target_revenue WHERE time_bill_dt_date = '12/31/2013';</pre>	
58.		<pre>DELETE FROM target_revenue WHERE time_bill_dt_date = '12/31/2013'; Commit;</pre>	To test the feature further, let us start by deleting data on the target table for 31-DEC-2013
59.			Switch to your Oracle Analytics page
60.		-	As we re-run the Data Flow, no data should be loaded, because we have not incremented our source table at all
61.		<ul style="list-style-type: none"> Click on Run Data Flow 	
62.		-	Inspect your Data Flow to make sure it's completed
63.		-	Switch to SQL Developer

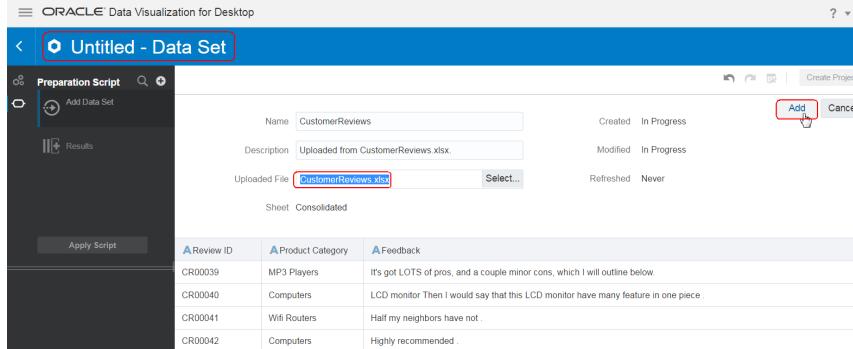
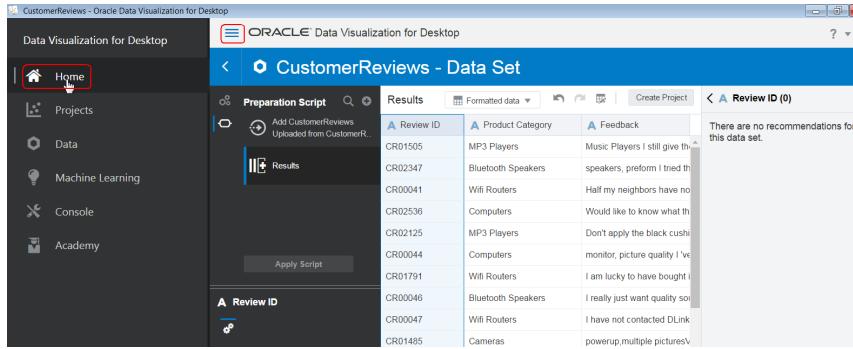
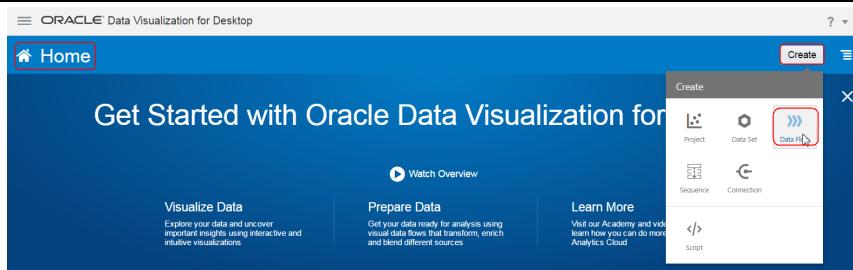
Step	View	Click stream	Talk stream
64.		<pre>SELECT MIN(time_bill_dt_date), MAX(time_bill_dt_date) FROM TARGET_REVENUE;</pre>	<p>Let's go back to SQL Developer and check again Target table.</p> <p>The Data Flow execution did not insert any records into the target table and hence no data for '29-DEC-2013' is available in the target table.</p>
65.		-	<p>Let's now run a second test: manually insert new records in the source table. Let us load records for '01-JAN-2014' into the FACT_REVENUE table</p>
66.		<pre>CREATE TABLE FACT_REVENUE_INC AS SELECT * FROM FACT_REVENUE where 1=0;</pre>	<p>Open DemoDB_BICS_SAMPLEAPP connection and create new blank table FACT_REVENUE_INC</p>
67.			<p>Insert new records from INSERT_INTO_FACT_REVENUE_INC.txt which has only records from '01-JAN-2014'</p>
68.			<p>Check the records</p>

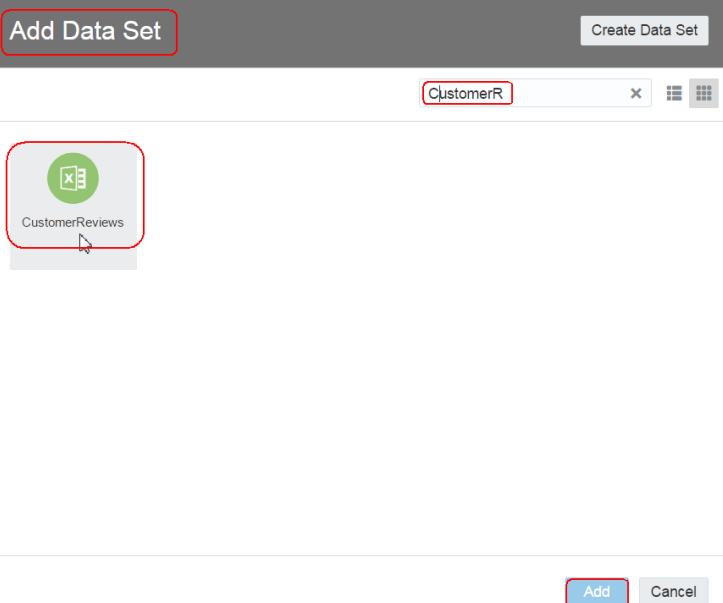
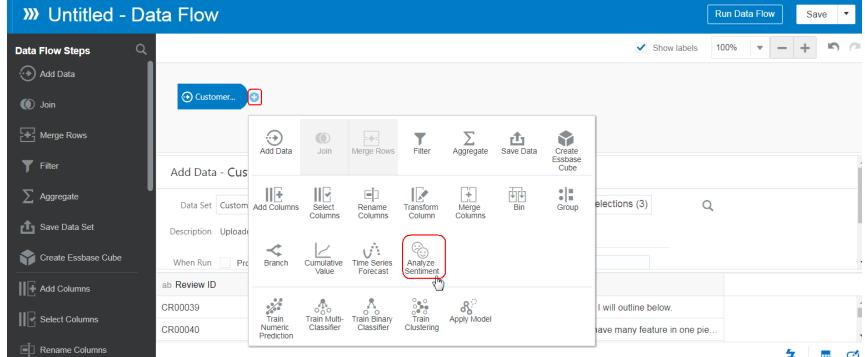
Step	View	Click stream	Talk stream
69.		INSERT INTO FACT_REVENUE SELECT * FROM FACT_REVENUE_INC; COMMIT;	Now you are going to insert the new rows from table FACT_REVENUE_INC to FACT_REVENUE
70.		SELECT MIN(time_bill_dt_date), MAX(time_bill_dt_date) FROM FACT_REVENUE;	Check the results and notice that we have new records for '01-JAN-2014' in our source table FACT_REVENUE
71.		-	Let's switch back to Oracle Analytics
72.		-	instance and run again the DF Incremental Data Flow
73.		-	
74.		-	Let's go back to SQL Developer

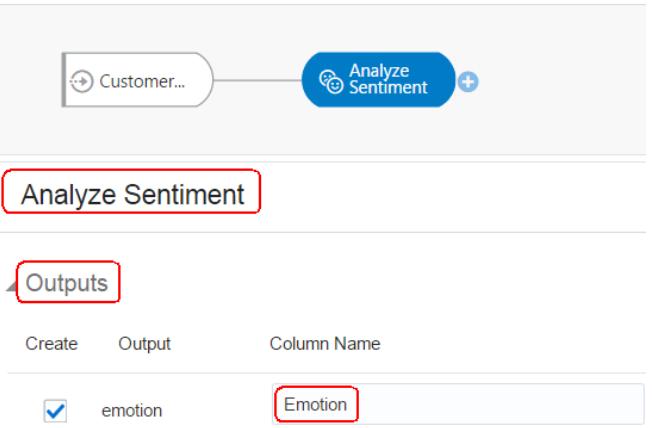
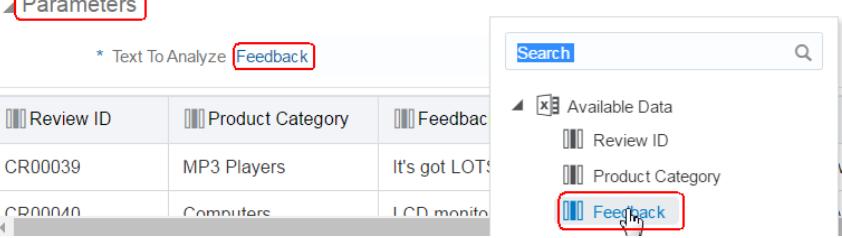
Step	View	Click stream	Talk stream
75.	 <pre> SELECT MIN(time_bill_dt_date), MAX(time_bill_dt_date) FROM TARGET_REVENUE; </pre>	<pre> SELECT MIN(time_bill_dt_date), MAX(time_bill_dt_date) FROM TARGET_REVENUE; </pre>	<p>and check again Target_Revenue table.</p> <p>Once completed, the data range in the target table shows that records have been incremented with '01-JAN-14' day.</p>
76.		-	<p>In this section you have learnt about Incremental Load feature of the Data Flows which is available with Database sources in Analytics Cloud.</p> <p>You defined in the source dataset a key column to be used to indicate new data (e.g. TIME_BILL_DT_DATE) since the last run. That's enable the Data Flow execution of the with only the new data, appends data rather than replace it.</p>

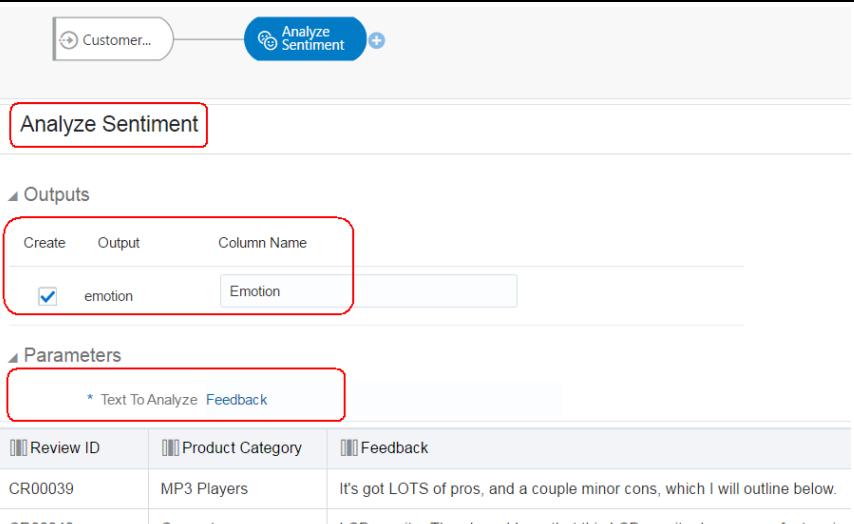
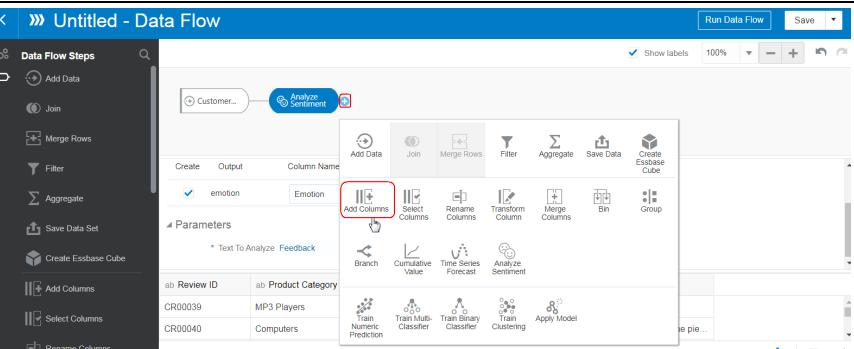
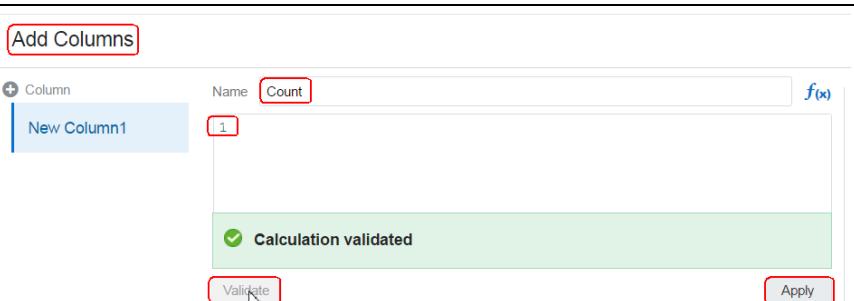
8. Create Sentiment Analysis and Enrich Data Set with Data Flow Step

Step	View	Click stream	Talk stream
1.			<p>In this demo, you will analyze a series of customer reviews, to understand the customer sentiment (emotions, feelings, likes, dislikes) regarding your products.</p> <p>Outcomes of a Sentiment Analysis can be categorized under Polarity (Negative, Neutral or Positive)</p>
2.			<p>Oracle Analytics has inbuilt capabilities to perform Sentiment Analysis on textual data. To invoke sentimental functionality, you add the twitter data set and create a data flow using the data set</p> <p>In Oracle Data Visualization, sentiment analysis is implemented using Python. To invoke it add Analyze Sentiment node to the dataflow.</p>
3.			<p>Let's create the dataset.</p> <p>Note: you have extensively created data sets during these lectures, so you get the main print screens as guidance, rather than step by step details</p>
4.			<p>From Home Page click Create > Dataset</p>

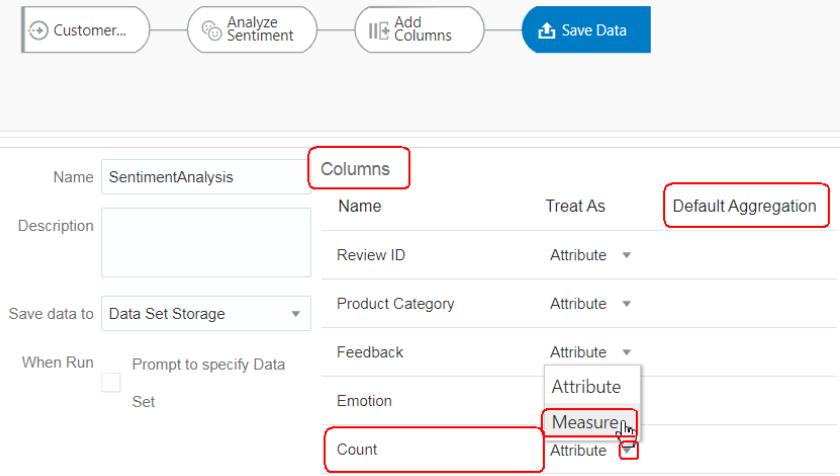
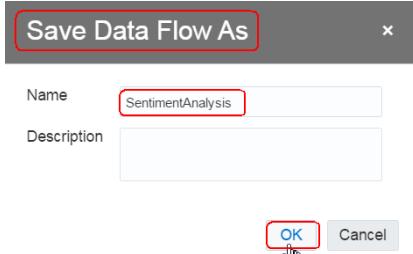
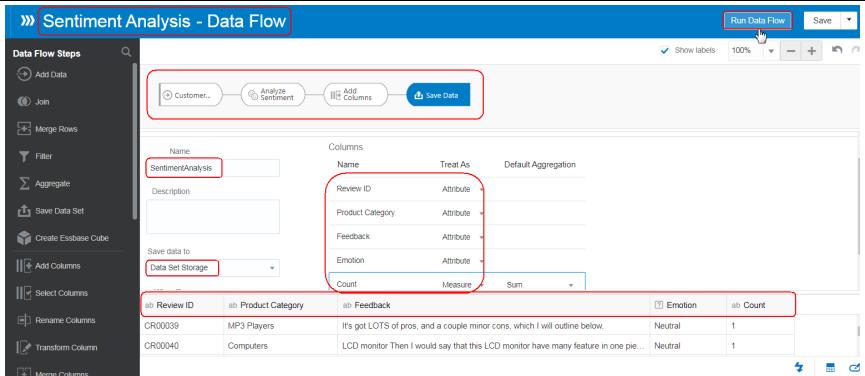
Step	View	Click stream	Talk stream
5.			Look for your CustomerReviews.xlsx file and click Add
6.			Go back to the Home Page
7.			Let's create our Data Flow
8.			From Home Page click Create > Data Flow

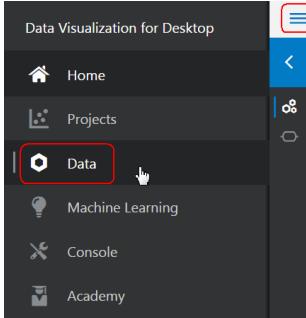
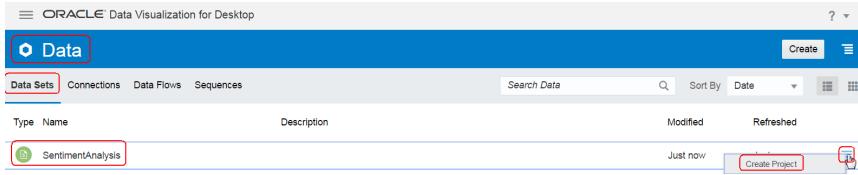
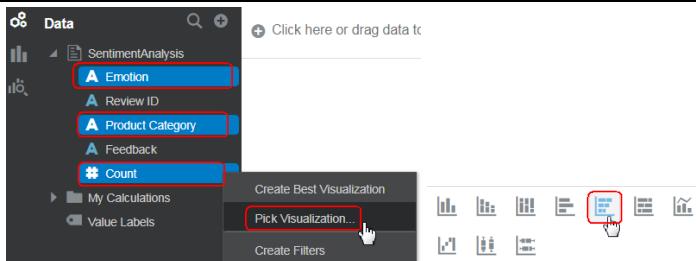
Step	View	Click stream	Talk stream
9.		<ul style="list-style-type: none"> - Select CustomerReviews - Click Add 	You choose CustomerReview Data Set that you have just created.
10.		<ul style="list-style-type: none"> - Click on Add Step  - Select Analyze Sentiment 	Add Step Analyze Sentiment

Step	View	Click stream	Talk stream									
11.	 <p>Analyze Sentiment</p> <p>Outputs</p> <p>Create Output Column Name</p> <p><input checked="" type="checkbox"/> emotion Emotion</p>	<ul style="list-style-type: none"> - Click on Outputs > Column Name and type Emotion 	<p>In the Analyze Sentiment pane, Output section, give a column name to capture the emotion value. Rename the default column as 'Emotion'</p>									
12.	 <p>Parameters</p> <p>* Text To Analyze Feedback</p> <table border="1"> <tr> <th>Review ID</th> <th>Product Category</th> <th>Feedback</th> </tr> <tr> <td>CR00039</td> <td>MP3 Players</td> <td>It's got LOT'S</td> </tr> <tr> <td>CR00040</td> <td>Computers</td> <td>LCD monitor</td> </tr> </table> <p>Available Data</p> <ul style="list-style-type: none"> Review ID Product Category Feedback 	Review ID	Product Category	Feedback	CR00039	MP3 Players	It's got LOT'S	CR00040	Computers	LCD monitor	<ul style="list-style-type: none"> - Click on Parameters > Text To Analyze and select Feedback 	<p>In the Parameters section, choose the column "Feedback" with text content to analyze</p>
Review ID	Product Category	Feedback										
CR00039	MP3 Players	It's got LOT'S										
CR00040	Computers	LCD monitor										

Step	View	Click stream	Talk stream									
13.	 <p>Analyze Sentiment</p> <p>Outputs</p> <p>Create Output Column Name</p> <p>emotion Emotion</p> <p>Parameters</p> <p>* Text To Analyze Feedback</p> <table border="1"> <thead> <tr> <th>Review ID</th> <th>Product Category</th> <th>Feedback</th> </tr> </thead> <tbody> <tr> <td>CR00039</td> <td>MP3 Players</td> <td>It's got LOTS of pros, and a couple minor cons, which I will outline below.</td> </tr> <tr> <td>CR00040</td> <td>Computers</td> <td>I CD monitor Then I would say that this LCD monitor have many feature in it.</td> </tr> </tbody> </table>	Review ID	Product Category	Feedback	CR00039	MP3 Players	It's got LOTS of pros, and a couple minor cons, which I will outline below.	CR00040	Computers	I CD monitor Then I would say that this LCD monitor have many feature in it.	-	-
Review ID	Product Category	Feedback										
CR00039	MP3 Players	It's got LOTS of pros, and a couple minor cons, which I will outline below.										
CR00040	Computers	I CD monitor Then I would say that this LCD monitor have many feature in it.										
14.	 <p>» Untitled - Data Flow</p> <p>Data Flow Steps</p> <ul style="list-style-type: none"> Add Data Join Merge Rows Filter Aggregate Save Data Set Create Essbase Cube Add Columns Select Columns Rename Columns <p>Run Data Flow Save</p> <p>Customer... Analyze Sentiment</p> <p>Create Output Column Name</p> <p>emotion Emotion</p> <p>Add Step +</p> <p>Parameters</p> <p>* Text To Analyze Feedback</p> <p>Review ID Product Category</p> <table border="1"> <thead> <tr> <th>Review ID</th> <th>Product Category</th> </tr> </thead> <tbody> <tr> <td>CR00039</td> <td>MP3 Players</td> </tr> <tr> <td>CR00040</td> <td>Computers</td> </tr> </tbody> </table> <p>Branch Cumulative Value Time Series Forecast Analyze Sentiment</p> <p>Train Numeric Prediction Train Multi Classifier Train Binary Classifier Train Clustering Train Quaterning Apply Model</p>	Review ID	Product Category	CR00039	MP3 Players	CR00040	Computers	<ul style="list-style-type: none"> - Click on Add Step - Select Add Column 	Let's Add a Count column			
Review ID	Product Category											
CR00039	MP3 Players											
CR00040	Computers											
15.	 <p>Add Columns</p> <p>Name Count</p> <p>New Column1</p> <p>Calculation validated</p> <p>Validate Apply</p>	<ul style="list-style-type: none"> - Click on Name and type Count - Type 1 as calculations - Click Validate - Click Apply 										

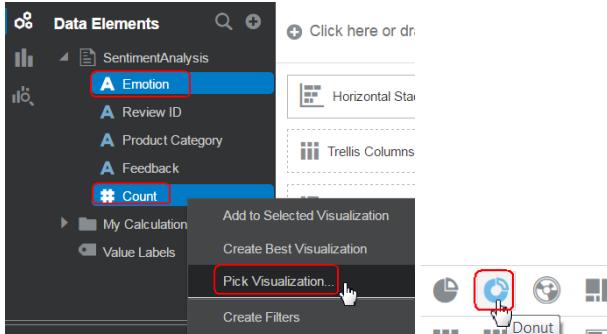
Step	View	Click stream	Talk stream
16.	<p>The screenshot shows the Oracle Analytics Data Flow interface. A flow is defined starting from a 'Customer...' step, followed by an 'Analyze Sentiment' step, and finally an 'Add Columns' step. The 'Add Columns' step has a red box around it. Below the flow, there is a table with three columns: 'Review ID', 'Product Category', and 'Feedback'. A new column 'Count' is being added with a value of 1. Another column 'Emotion' is also being added. The 'Add Columns' step has a red box around its 'Add Column' button.</p>	-	Check the two new columns
17.	<p>The screenshot shows the Oracle Analytics Data Flow interface. The 'Add Columns' step has a red box around its 'Add Step' button. Below the flow, there is a toolbar with several icons: 'Add Data', 'Join', 'Merge Rows', 'Filter', 'Aggregate', 'Save Data' (which has a red box around it), and 'Create Essbase Cube'.</p>	<ul style="list-style-type: none"> - Click on Add Step - Select Save Data 	Choose to Save Data
18.	<p>The screenshot shows the 'Save Data Set' dialog box. It has fields for 'Name' (set to 'SentimentAnalysis'), 'Description', and 'Save data to' (set to 'Data Set Storage'). The 'Name' and 'Save data to' fields have red boxes around them.</p>	<ul style="list-style-type: none"> - Type on Name: SentimentAnalysis 	Give the Data Set Name: SentimentAnalysis

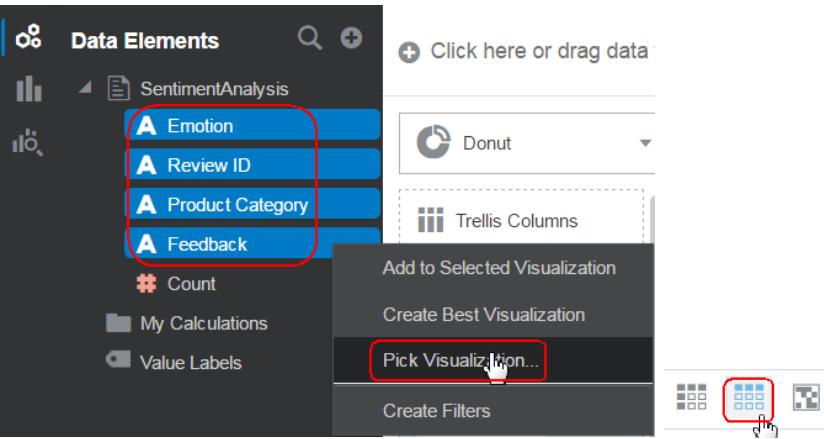
Step	View	Click stream	Talk stream
19.		-	Change the Default Aggregation for Emotions to Measure
20.		- Click Save	and save the dataflow
21.		- Type on Name: Sentiment Analysis - Click OK	
22.		- Click on Run Data Flow	Run SentimentAnalysis Data Flow

Step	View	Click stream	Talk stream
23.			Let's check the new created Data Set
24.			Go to Data
25.			Let's create a Project out of it
26.		<ul style="list-style-type: none"> - Select Sentiment Analysis Data Set - Click on Actions Menu - Click on Create Project 	
27.			<p>You are going to create a number of visualizations, for better understanding of the generated feedback for your product category.</p> <p>Note: We have been created visualizations in previous courses so we are not going to have a step by step approach. Please check the print screens for hints</p>
28.			Pick Horizontal Stacked vis

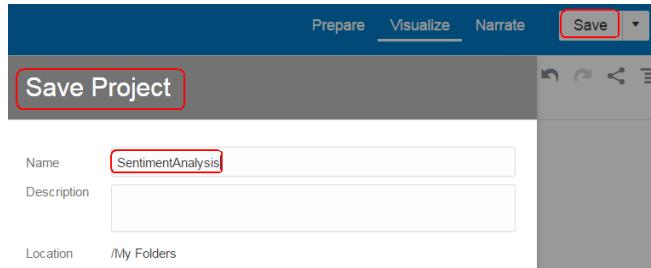
Step	View	Click stream	Talk stream
29.			You may want to switch Emotions with Product Category
30.			You also want to align colors with Sentiment , so Negative would be red colored and Positive been green.
31.			You start by choosing Neutral as yellow (#f7f37b)

Step	View	Click stream	Talk stream
32.			
33.			Positive as green (#68c182)
34.			And Negative as red (#ed6647)
35.			You have now colors aligned in the chart

Step	View	Click stream	Talk stream
36.			Let's have a count (number of reviews) by Sentiment
37.			Pick up Donut visualization
38.			You see a high number (2/3) of Emotions are Neutral
39.			Let's add the last visualization to check on Feedback narrative

Step	View	Click stream	Talk stream
40.			Pick up Table visualization
41.			<p>Use the print screen as a guidance.</p> <p>The report on the top left gives us the number of reviews by Product Category and Sentiment. Bluetooth Speakers category seems to have a very high percentage of Positive reviews and is the second largest category in terms of number of reviews.</p> <p>The top right report gives us a just the sentiment ratio w.r.t. the number of reviews.</p> <p>The report at the bottom is simply displaying a list of all reviews color coded by sentiment. You can start now to read full Feedback reviews and focus on the relevant ones.</p>

Step	View	Click stream	Talk stream
42.	<p>A screenshot of the Oracle Analytics Canvas interface. On the left, the Data Elements panel shows 'SentimentAnalysis' with 'Emotion', 'Review ID', 'Product Category', 'Feedback', and 'Count' selected. Below that are 'My Calculations' and 'Value Labels'. In the center, a Pivot Table visualization titled 'Count by Emotion, Product Category' is displayed. The table has columns for 'Negative Count', 'Neutral Count', 'Positive Count', and 'Total'. The data includes rows for Bluetooth Speakers, Cameras, Computers, MP3 Players, and WiFi Routers, with a 'Total' row at the bottom. A context menu is open over the table, with options like 'Duplicate Visualization', 'Edit', 'Copy Visualization', 'Copy All Data', 'Hide Value Labels', and 'Delete Visualization'. The bottom of the screen shows 'Canvas 1' and 'Canvas 2' tabs.</p>		<p>You add a new Canvas with Pivot Table with total number of comments by Emotion and Product Category.</p>
43.	<p>The same screenshot as Step 42, but the 'Edit' option in the context menu is highlighted with a red box. The menu items are: Drill to Attribute/Hierarchy..., Use as Filter, Add Statistics, Color, Edit, Hide Value Labels, Copy Visualization, Copy All Data, and Delete Visualization.</p>		<p>Duplicate the visualization</p>
44.	<p>The same screenshot as Step 42, but the 'Display As' setting in the Data Elements panel is changed to 'Percent Of Row Values'. The table now shows percentages instead of raw counts. The last row shows the total percentages for each column: Negative (2.97%), Neutral (72.34%), Positive (24.69%), and Total (100.00%).</p>		<p>And choose to display as percent of Row Values. MP3 Players have the highest percentage of Negative reviews among Product Category</p>

Step	View	Click stream	Talk stream
45.	 <p>Save Project</p> <p>Name: SentimentAnalysis</p> <p>Description:</p> <p>Location: /My Folders</p> <p>Save Cancel</p>		<p>Save the project as SentimentAnalysis</p>
46.			<p>Summary: In just a few minutes, you were able to:</p> <ul style="list-style-type: none"> - Analyze the customer feedback and display it as color coded by sentiment - Further understand the sentiment by the product category

# A Vehicle Id	# A Accident Id	# A Person Id	# A Fatality Id
	#		
General	General	General	General
Name	A Vehicle Id	Name	A Accident Id
Treat As	Measure	Treat As	Measure
Data Type	Text	Data Type	Number
Aggregation	Count Distinct	Aggregation	Count

A Event Month	A Event Year	P Age
#	#	
General	General	General
Name	Event Month	Name
Treat As	Attribute	Treat As
Data Type	Number	Data Type
		Number
		Aggregation
		Average