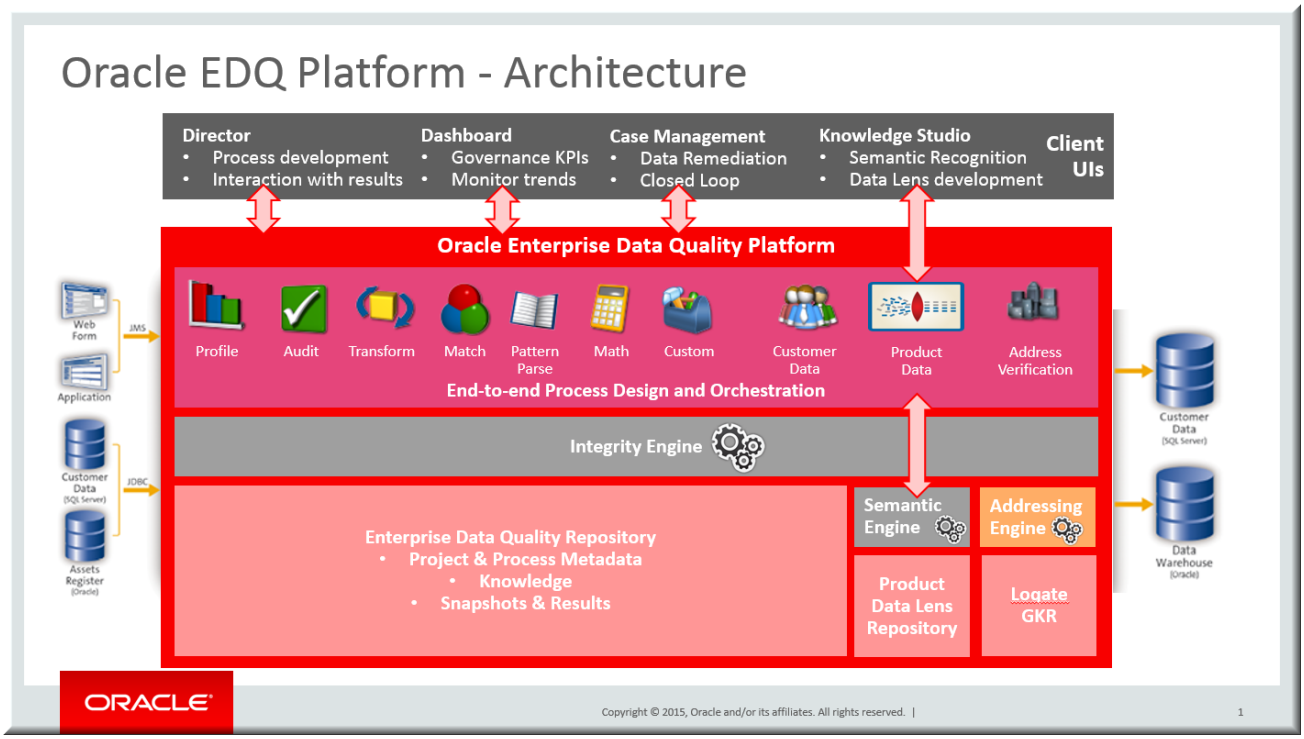
Data Integration Platform Cloud
Governance Edition

Data Quality Hands On Lab

INTRODUCTION:

Enterprise Data Quality is Oracle's premiere solution for delivering 'Data Fit for Use' and 'Accurate' Analytics that satisfy an organization's current and emerging Data Governance requirements for your Data Warehouse / Data Mart / Data Lake environments/initiatives. Oracle Enterprise Data Quality provides organizations with an integrated suite of data quality tools that provide an end-to-end solution to measure, improve and manage the quality of data from any domain, including Customer / Citizen / Student / Employee / Patient and others. Oracle Enterprise Data Quality also combines powerful data profiling, cleansing, matching and monitoring capabilities while offering unparalleled ease of use.

Oracle Enterprise Data Quality Component Architecture



About the Lab:

During the lab, we'll covering data profiling, data auditing & data enrichment topics. We have a sample data named as 'Mortgage EDQ Sample'. We'll be working on this data set.

In the Lab 1a; you'll get familiar with EDQ, you'll import your sample data into EDQ repository and you'll create your project where we'll run our tasks.

In the Lab 1b; you'll discover the data relationships, anomalies, standardization variances and resulting negative impacts on data by using some of the advanced data profiling functionalities.

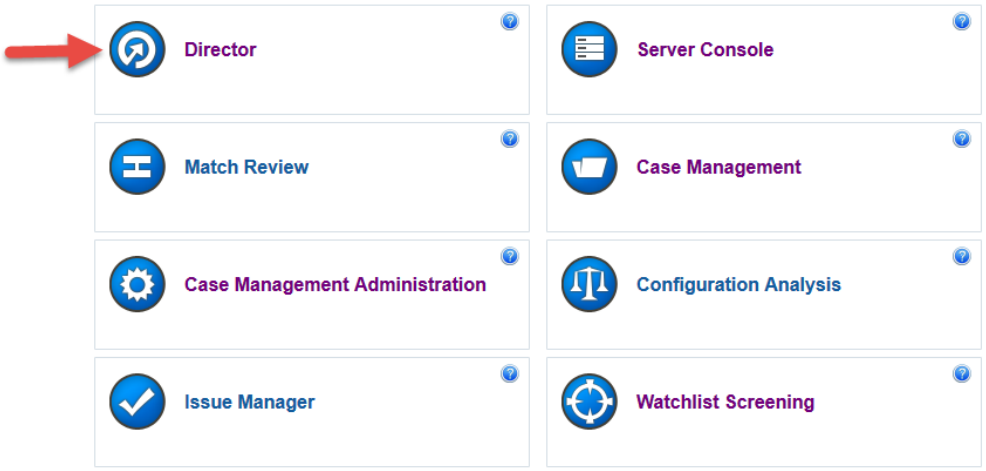
In the Lab 2; we'll expand upon our Data Profiling activities performed in Lab 1 and carry out Audit data checks using EDQ Audit processors.

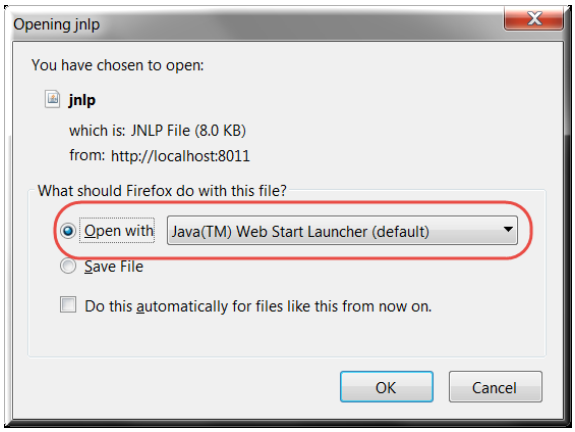
Lab 1a: Enterprise Data Quality Director

Navigate to the Enterprise Data Quality Launchpad and start Director

1. Launch an internet browser (Mozilla Firefox, Internet Explorer, Google Chrome) and navigate to URL provided in 'Access Details' document shared with you.

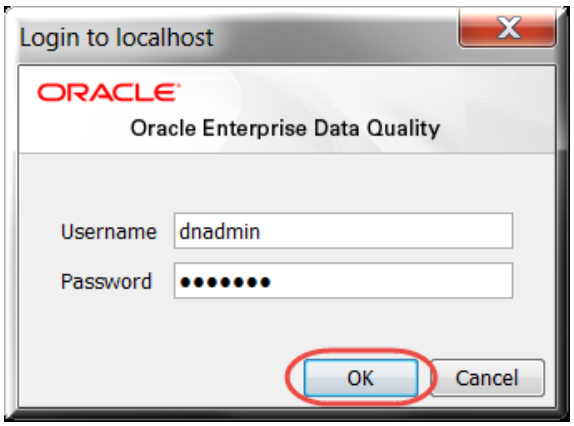
2. Click on the Director icon to launch the Director user interface.



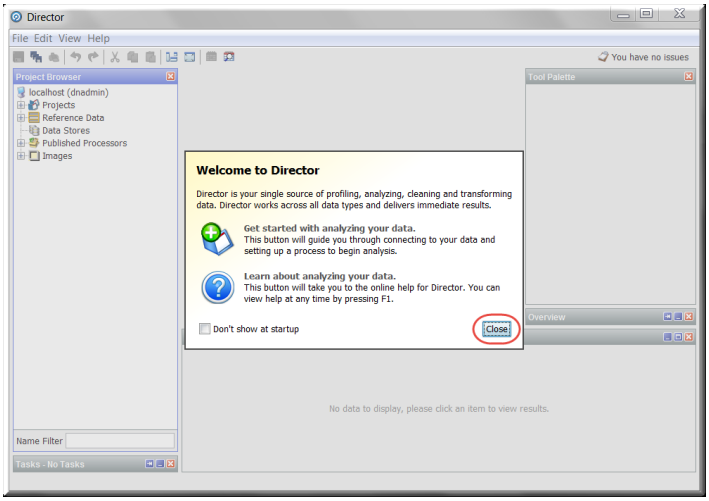


3. You may be prompted to open a file. Choose open with and select the Java Web Start application and click **OK.** If you are prompted to allow the application to run, click **Run.**

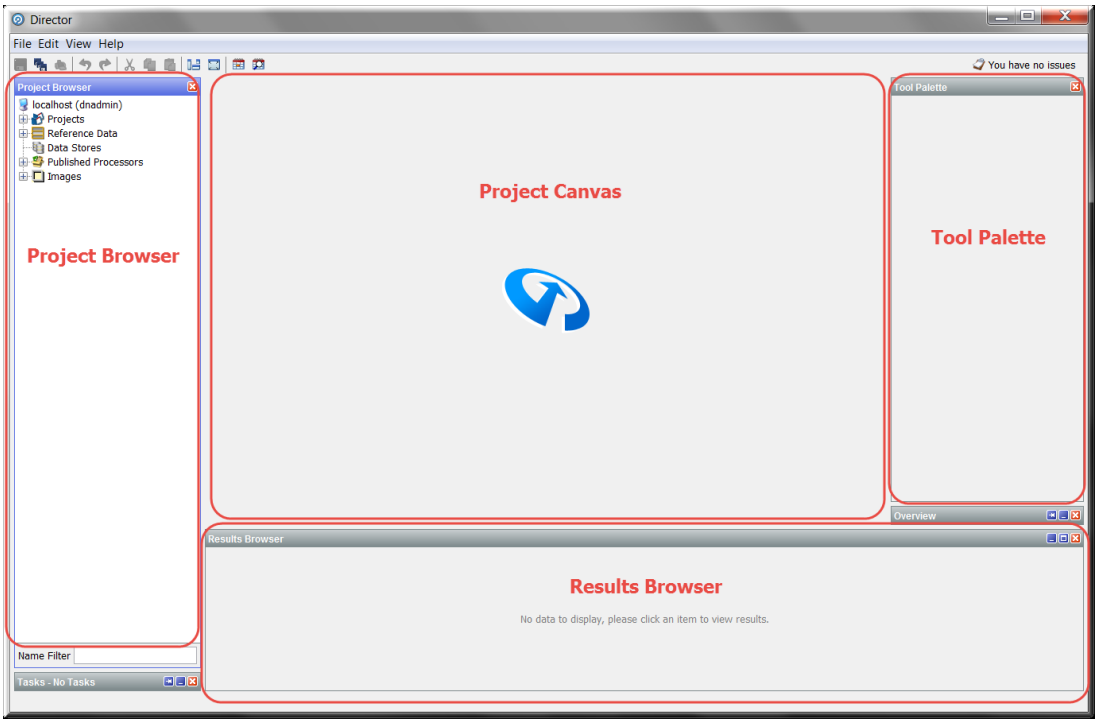
4. You will be prompted to login to Enterprise Data Quality. Use the credentials provided in 'Access Details' document shared with you.



a. After Director launches, click on **close** in the **Welcome to Director** message:



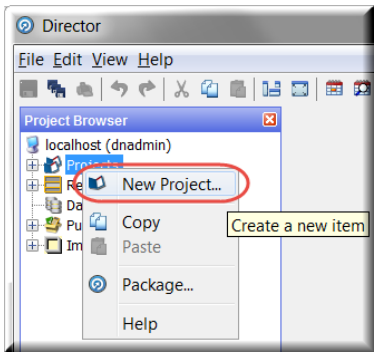
5. Take a moment to familiarize yourself with the Terminology of each of the four different areas of the **Director** application.



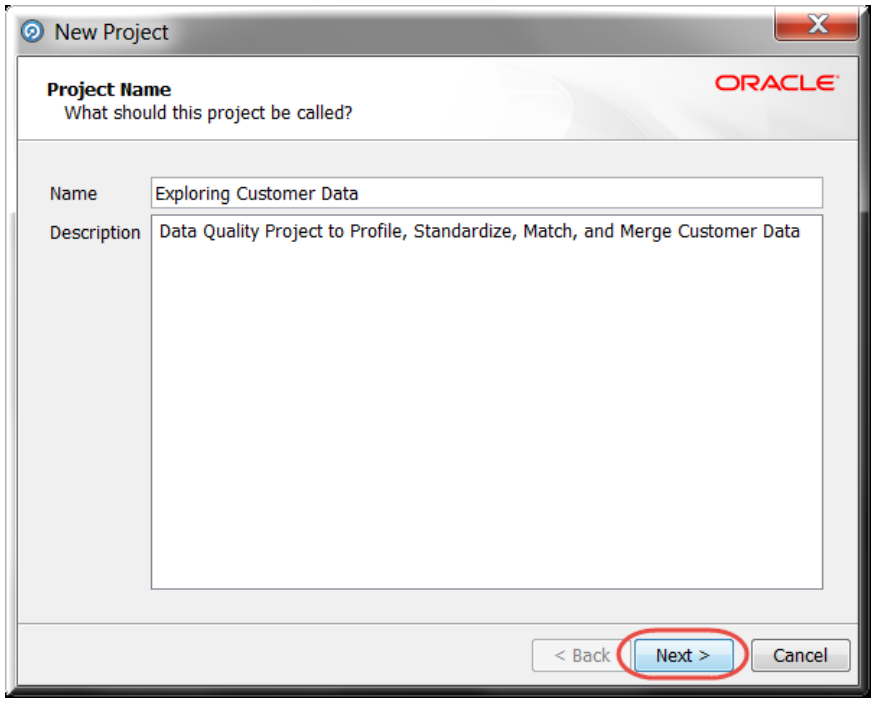
Create a New Project

Projects are created in the Project Brower and are generally utilized to hold data and processes related to a Data Quality initiative. You can set permissions and access levels at a project level. We will now begin a data quality project utilizing sample customer data from a US based order management system throughout the rest of the lab.

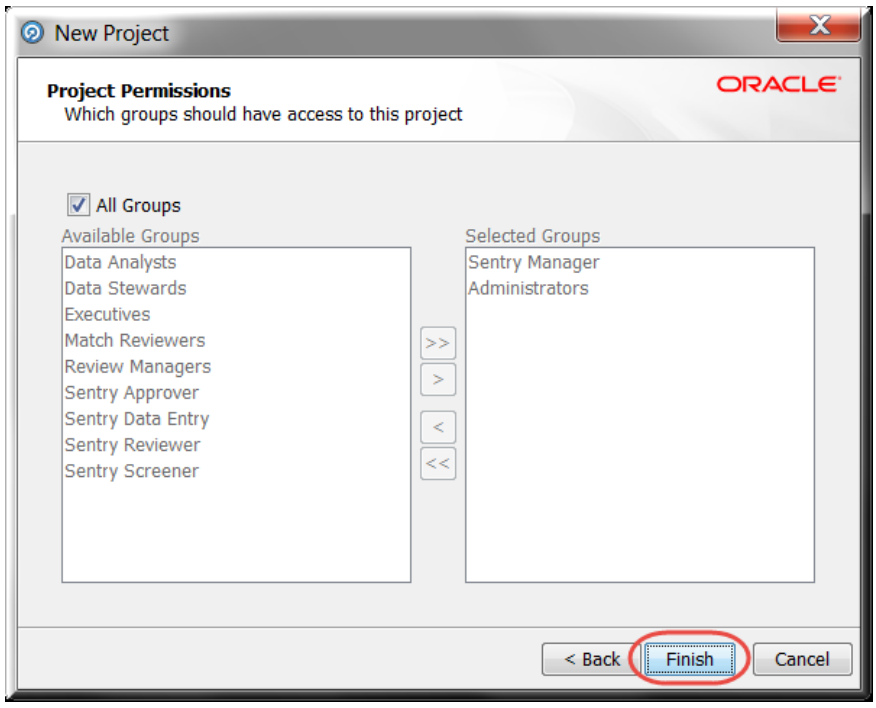
1. In the **Project Browser**, right-click **Projects** and select **New Project…** to start the Wizard.



2. For **Name**, enter *Project User3* and optionally add a **Description**. Click **Next >** to continue



3. Ensure the **All Groups** checkbox is selected in **Project Permissions**. This will ensure any user can view and use the project. Click **Finish** to create the new project



Add a Data Store

Now that we have created a Project for the Labs, the next step is enabling access to your Data that needs Profiling / Enrichment / Standardizing and optional Match / Merge / De-duplication. Turning your Data into:

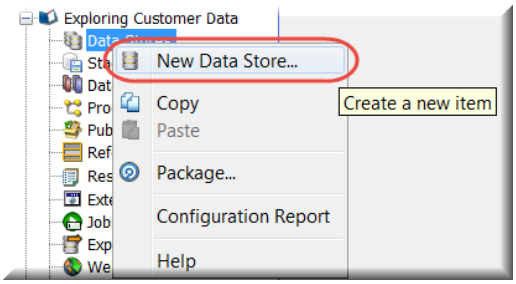
- ‘not just Data’, but ‘Data Fit for Use’
- ‘not just Analytics’, but ‘Accurate Analytics’

A Data Store is a connection to a store of data, whether the data is stored in a database or in one or more files. The data store may be used as the source of data for a process, or you may export written Staged Data results of a process to a data store, or both.

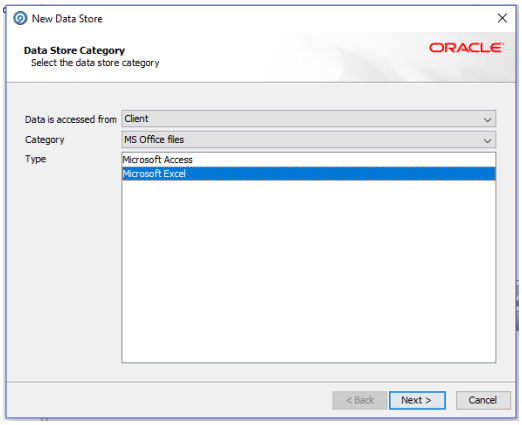
EDQ supports native connections to many types of type of data stores. For our lab, we’ll be using an excel file as the data source.

Take a moment to familiarize yourself with different source & data types supported by EDQ.

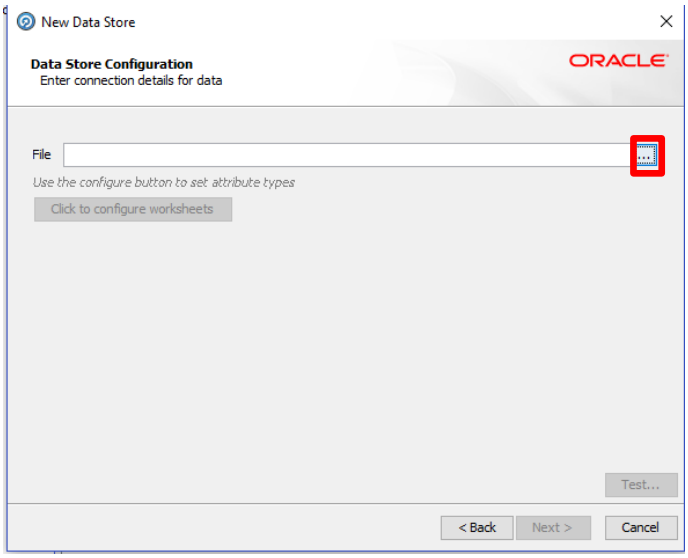
1. Expand the newly created project **Project User3**, right click **Data Stores**, and select **New Data Store** to launch the **New Data Store** Wizard



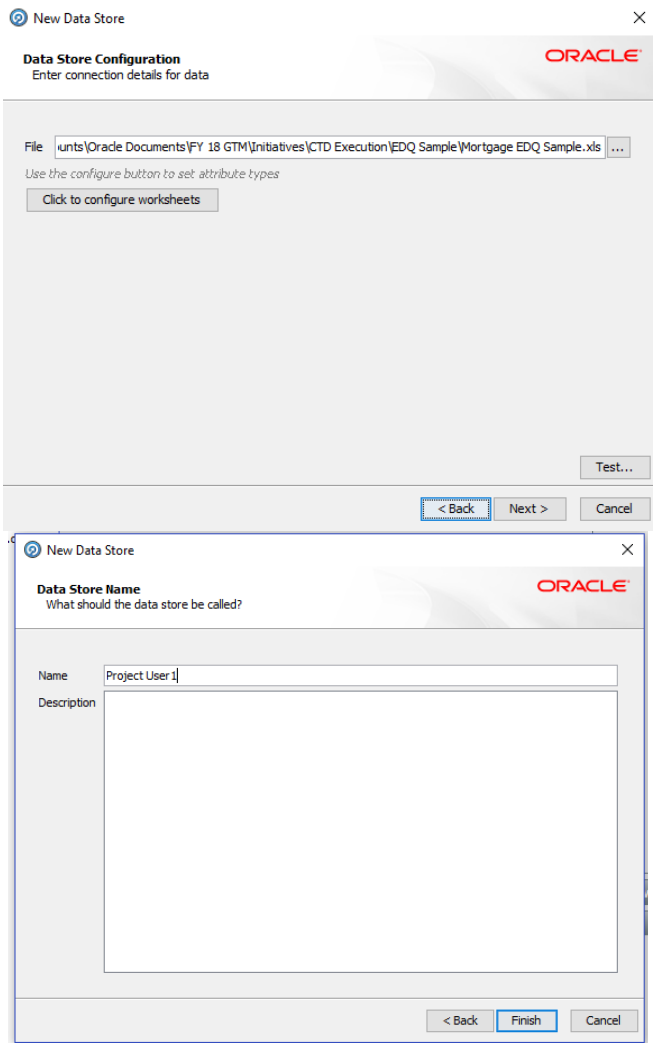
2. Select **Client** from ‘Data is accessed from’ and **MS Office Files** from ‘Category’, then select **Microsoft Excel**. Click **Next** > to continue.



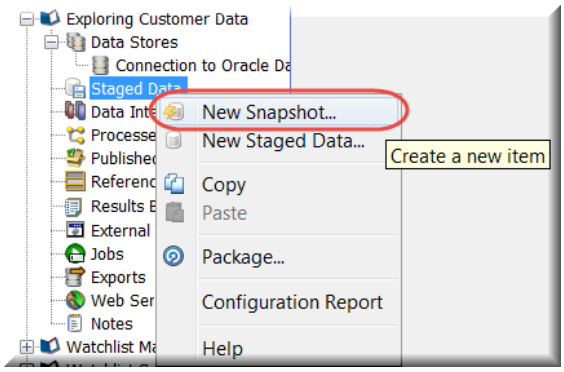
3. Select '**Mortgage EDQ Sample**' file provided to you.



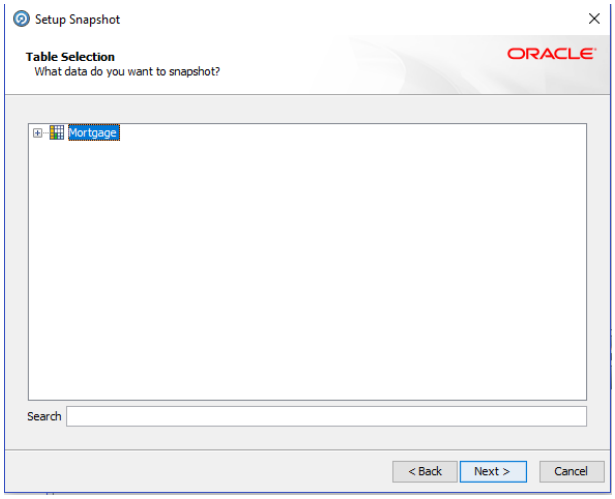
4. Click **Next** and **Finish.** Give the name as **Data Store User3.**



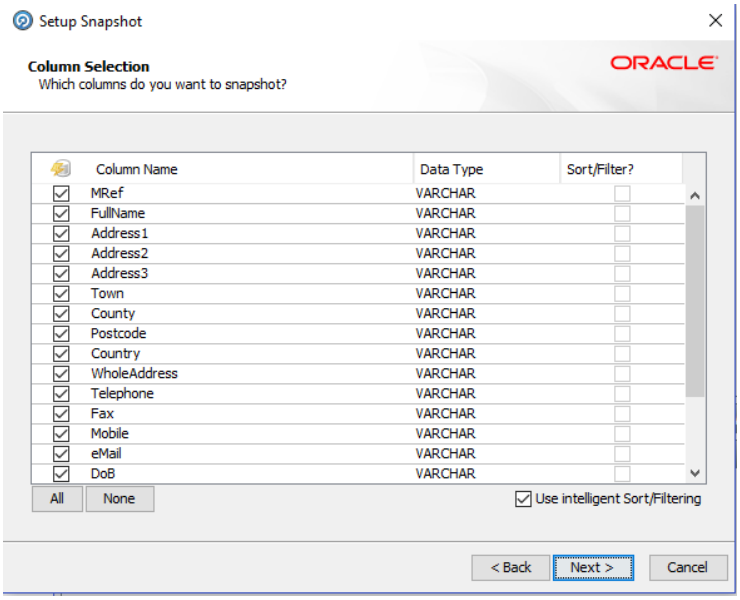
5. Navigate back to the **Project Browser** and right click **Staged Data** under your **Project User3** project and select **New Snapshot…**



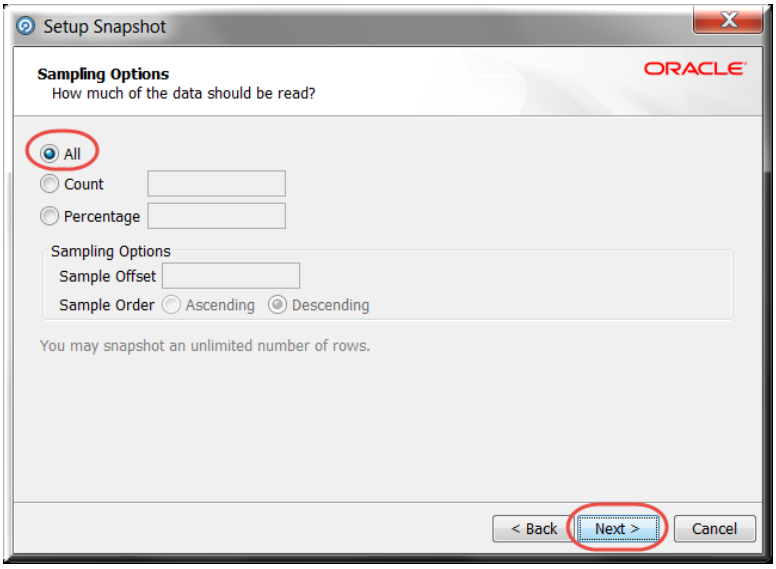
6. Select newly created **Data Store User3** then click **Next** to continue. This is where the data for the snapshot will come from. Keep Mortgage selected and click **Next.**



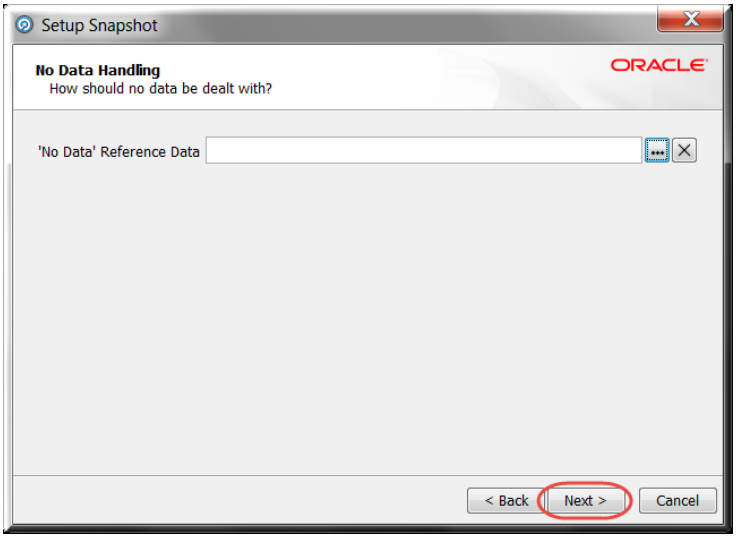
7. For **Column Selection**, ensure all columns are selected for setting up this data snapshot, then click **Next** > to continue.



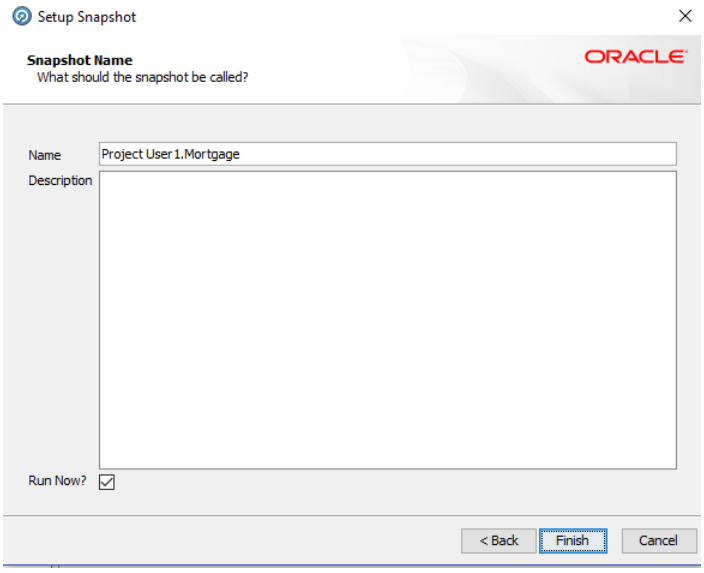
8. **Sampling Options** allows the behavior of the amount of data read into the snap shot to vary. By default, the **All** radio button selection is selected. If needed, you can specify a certain **Count** or **Percentage** of data to be read for the snapshot. In this example, select **All** for the sampling options, then click **Next** > to continue.



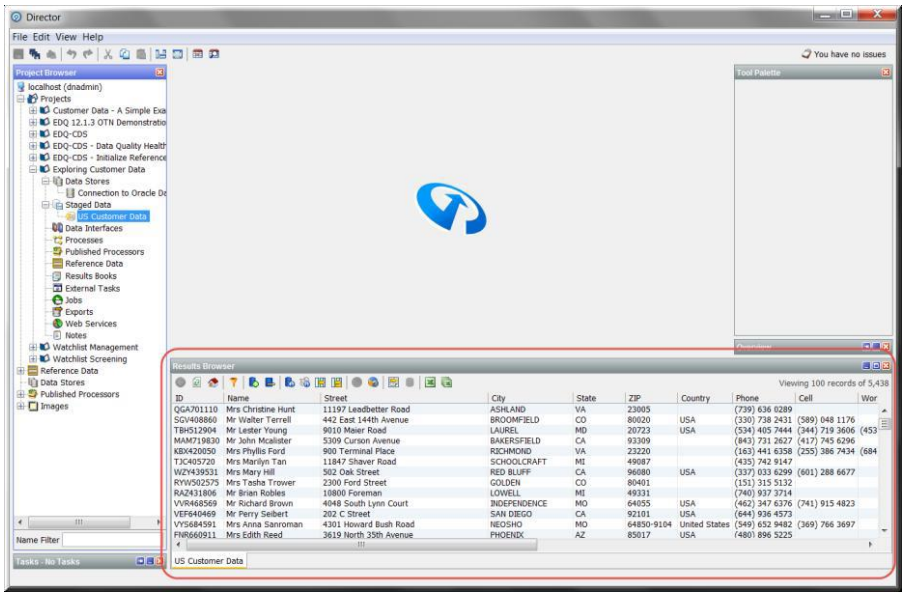
9. Leave the default empty value for the **'No Data' Reference Data** field. We will work with Reference Data later in another lab.



10. Give your snapshot the following name – **Staged Data User3.Mortgage**. Ensure the **Run Now** Checkbox is checked, then click **Finish** to complete and close the **New Snapshot** wizard.



Notice that after a short delay, the **Results Browser** is populated with data originating from the Oracle Database and sourced from an EDQ Snapshot. Taking the Snapshot causes Enterprise Data Quality to stage the data from the database into the EDQ data repository – meaning that a copy of the data from the database is placed in the Enterprise Data Quality repository. From now on we will be working with the data residing in the US Customer Data Snapshot.



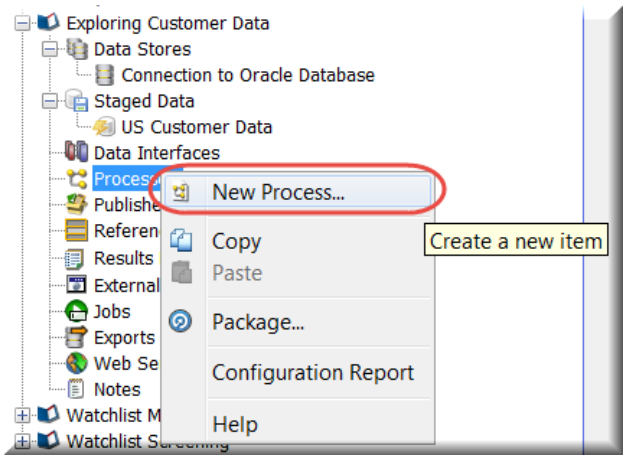
Lab 1b: Profiling your Data

The first step in improving the quality of your data is to understand it. You know you have 'Data' – but is it 'Data Fit for Use'? Enterprise Data Quality allows the user to quickly assess, find, investigate and understand anomalies regarding data content, standardization, relationships and duplication among others. EDQ enables users to understand their data by discovering, highlighting and communicating data anomalies within the data being profiled.

As you will learn, Profiling can lead to many different insights on your data sources and targets including outliers, minimum and maximum values, invalid dates and record completeness. It can also show the frequency with which particular values occur. For instance, how many unique values do you have in a field, and how many times is a particular value duplicated? These profiling topics and more are the focus of our following lab: Create and Run a Profiling Process.

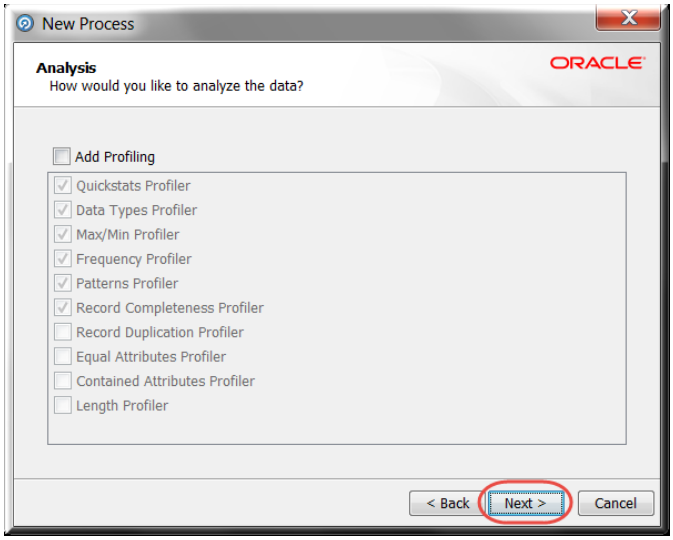
Create & Run A Profile Process

1. Navigate to the Project Browser and right-click on **Processes** under your **Project User3** Project, then click **New Process**…

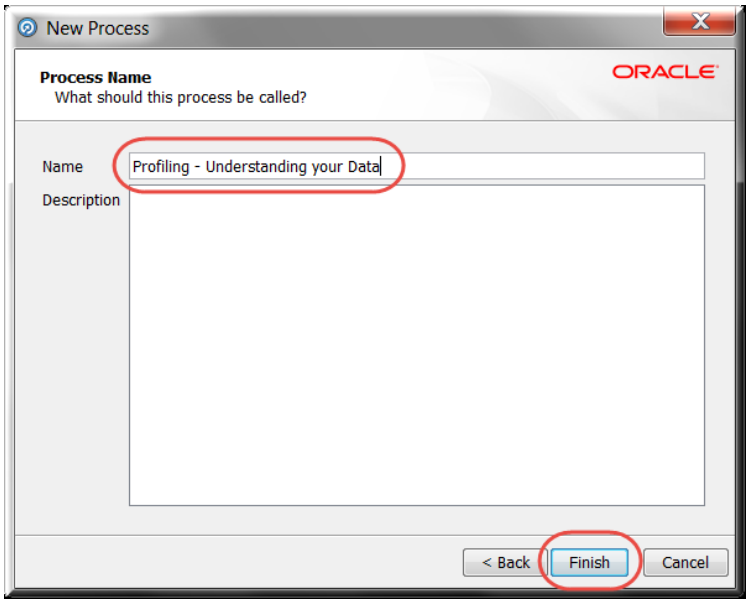


2. Select the previously Staged Data, **Staged Data User3.Mortgage**, then click **Next** > to continue.

3. Notice that you can optionally select **Add Profiling** while creating this New Process. We will add our own Profiling processors in the next few steps. Leave the checkbox unchecked and click **Next** > to continue.



4. Give this process a name: **Profiling – Understanding your Data** – then click Finish

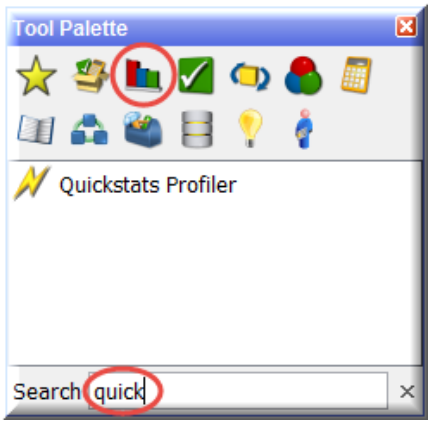


Congratulations! A tabbed Project Canvas is now presented with your newly created Process. You will note a **Reader** processor is automatically added to the Project Canvas.

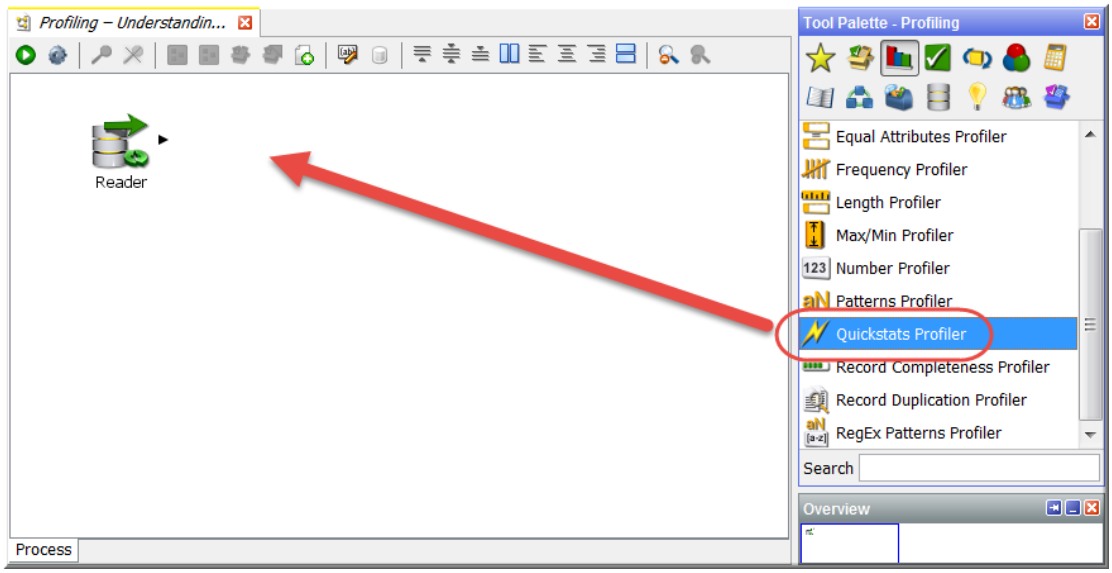
The term processor will be used to refer to the different pre-built objects that are dragged and dropped from the **Tool Palette**. In short, each processor can be configured in a process to perform some kind of operation on your data. In this case, since the Staged Data, US Customer Data, was selected while creating the process, a **Reader** processor that ingests **Mortgage Data** has already been added to the canvas.

5. Navigate to the **Tool Palette** and find the **Profiling** icon. Next, find and select the **Quickstats Profiler** among the Profiling processor family.

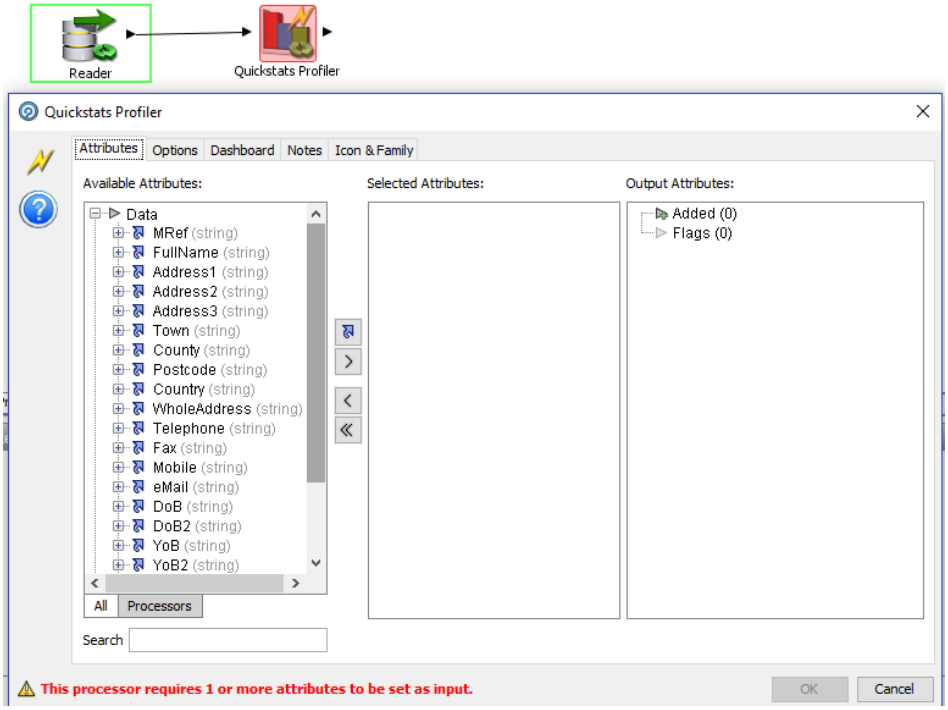
   *** Using the **Search** box underneath the Tool Palette allows you to quickly find Processors also.*



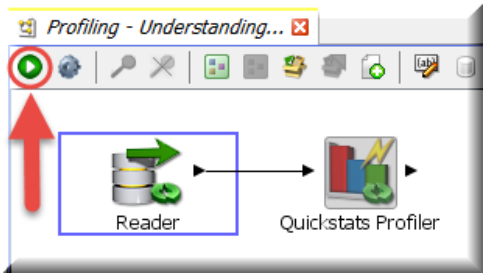
6. Drag and drop the **Quickstats Profiler** onto the Project Canvas



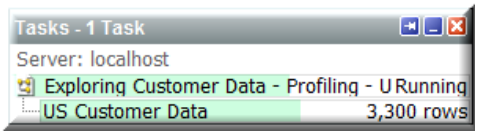
7. Hover over the output triangle of the **Reader** processor. An information tool-tip appears with the name and brief description of the processor. Click and drag from the output triangle of the **Reader** processor to the input triangle of the **Quickstats Profiler**. Upon successful connection and release of the mouse, the **Quickstats Profiler** configuration dialog will appear:



8. Note the message in red informing you that **This processor requires at least one attribute to be set as an input**. Click **the Select All** icon , as shown in the screenshot above. This will select all available attributes from our **Staged Data User3** staged data. Click **OK** to save.
9. The process now has a **Reader** and a **Quickstats Profiler**. Click the **Run** icon in the toolbar to run the process.

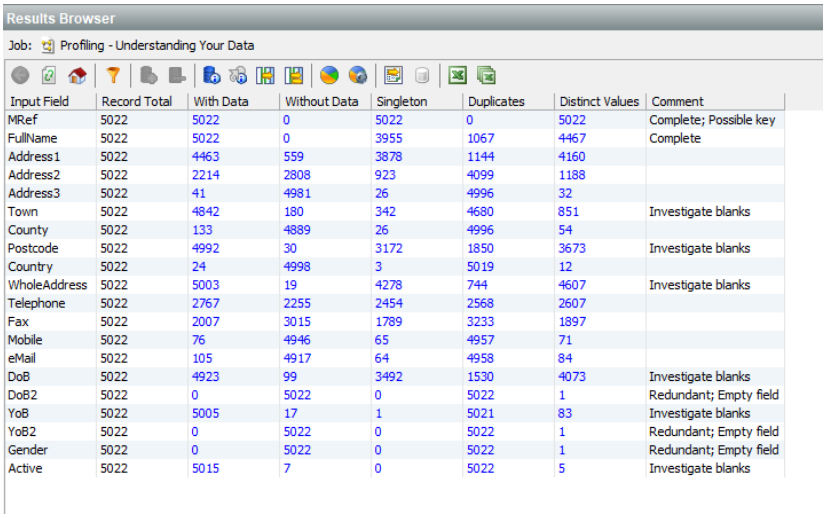


*** The progress can be observed in the **Task Bar** in the bottom-left of the Director as the process runs. When the process has finished, the 'not yet run' icons will disappear from the canvas to show that the processors have data associated with them.*



10. Click the **Reader** processor to see the raw input data stored in the staged data snapshot. This will be displayed in the **Results Browser**

11. Next click the **Quickstats Profiler** to see the output of the processor



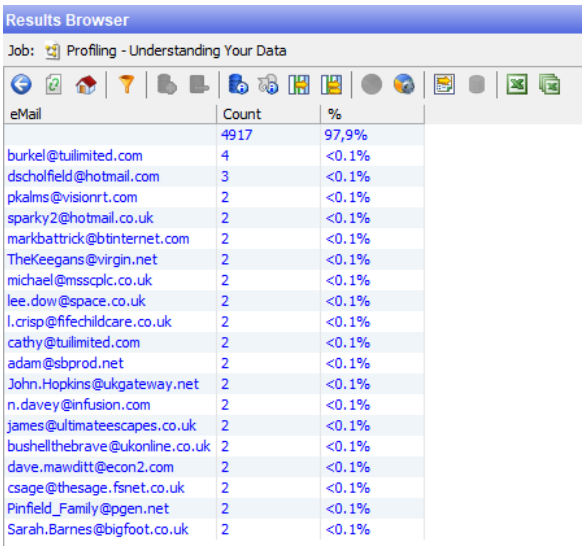
Input Field	Record Total	With Data	Without Data	Singleton	Duplicates	Distinct Values	Comment
MRef	5022	5022	0	5022	0	5022	Complete; Possible key
FullName	5022	5022	0	3955	1067	4467	Complete
Address1	5022	4463	559	3878	1144	4160	
Address2	5022	2214	2808	923	4099	1188	
Address3	5022	41	4981	26	4996	32	
Town	5022	4842	180	342	4680	851	Investigate blanks
County	5022	133	4889	26	4996	54	
Postcode	5022	4992	30	3172	1850	3673	Investigate blanks
Country	5022	24	4998	3	5019	12	
WholeAddress	5022	5003	19	4278	744	4607	Investigate blanks
Telephone	5022	2767	2255	2454	2568	2607	
Fax	5022	2007	3015	1789	3233	1897	
Mobile	5022	76	4946	65	4957	71	
eMail	5022	105	4917	64	4958	84	
DoB	5022	4923	99	3492	1530	4073	Investigate blanks
DoB2	5022	0	5022	0	5022	1	Redundant; Empty field
YoB	5022	5005	17	1	5021	83	Investigate blanks
YoB2	5022	0	5022	0	5022	1	Redundant; Empty field
Gender	5022	0	5022	0	5022	1	Redundant; Empty field
Active	5022	5015	7	0	5022	5	Investigate blanks

The **Quickstats Profiler** provides fundamental quality metrics for a number of records or transactions, highlighting:

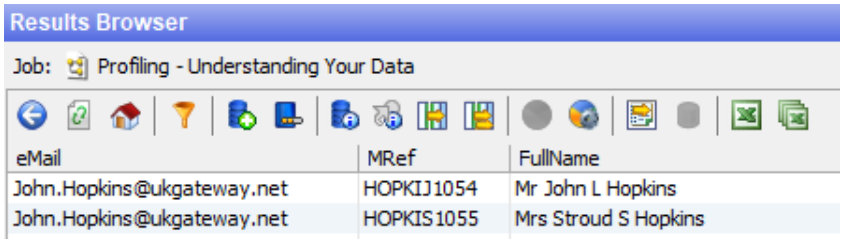
- Candidate key columns
- Completeness and missing data
- Duplication
- Uniqueness and diversity of values

For each **Input Field**, the number of records (**Record Total**), **With Data**, **Without Data**, **Singleton, Duplicates**, and **Distinct Values** are shown. These results can be observed and investigated to quickly find data anomalies. For instance, there are **4 Distinct Values** for the Gender attribute, when there should really only be two: Male and Female. You can also drill down on any blue text to see the data underneath.

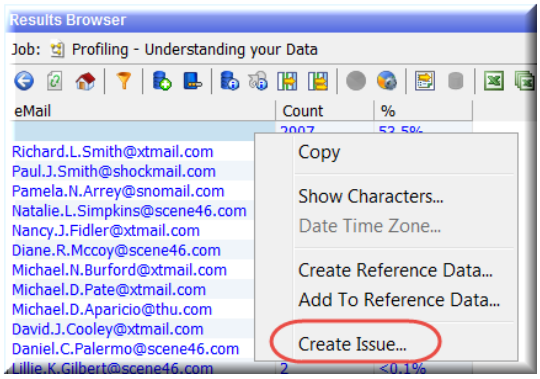
12. Click the number **4969** listed for **eMail** under the **Duplicates** column in the **Results Browser**. Click the **Count** hyperlink for the **eMail** address containing no content / blank.



13. Click  in the **Results Browser** to return to the previous view of Drill-down on one of the non-null (with data) values. We observe that there are a number of duplicate **eMail** values (**Count** of **2**) in the system that may require further investigation from a duplicate record standpoint

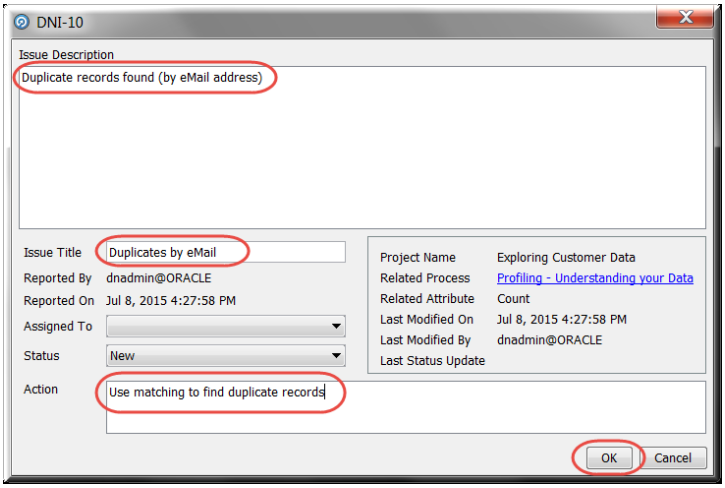


14. Enterprise Data Quality has many features to create a collaborative environment. To add an issue for duplicate (**Count** > 1) eMail records, right-click on a hyperlink field where the **Count** value **is 2** in the **Results Browser** and choose **Create Issue**…

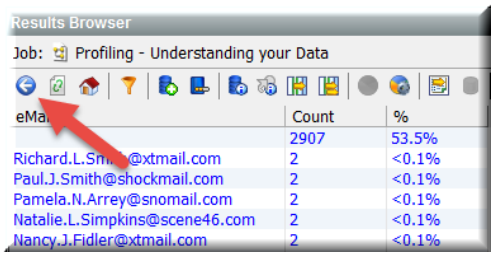


15. It is possible to assign the issue to yourself (**dnadmin@ORACLE**) or another user. You can also type in a follow-up **Action**: *Use matching to find duplicate records*. The issue

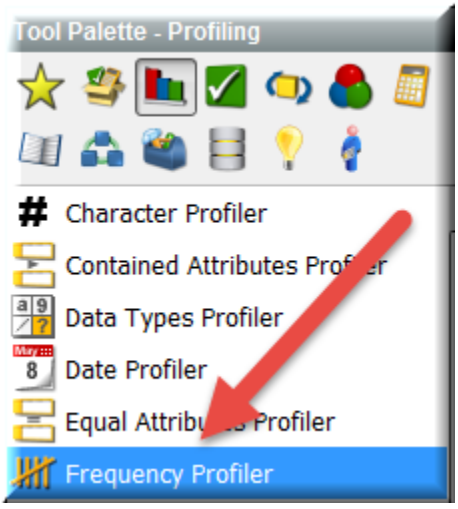
also includes a link to the process and results view where the issue was created. Under **Issue Description**, type *Duplicate records found (by eMail address).* Click OK and the issue is created



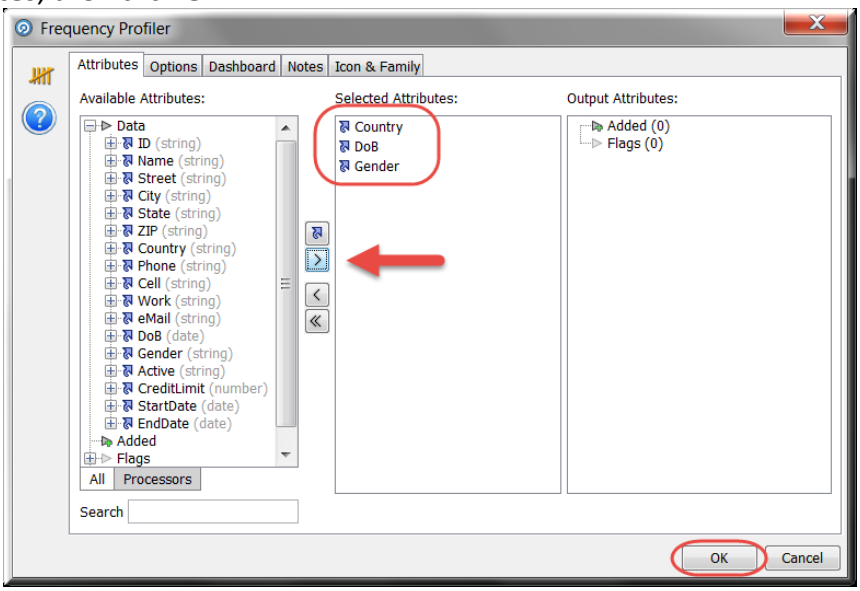
16. Click the back icon in the **Results Browser** to return to the results of the **Quickstats Profiler**



17. Return to the **Tool Palette - Profiling** and find the **Frequency Profiler**. Drag and drop the processor onto the Project Canvas and link the output triangle of the **Quickstats Profiler** to the input of the **Frequency Profiler**



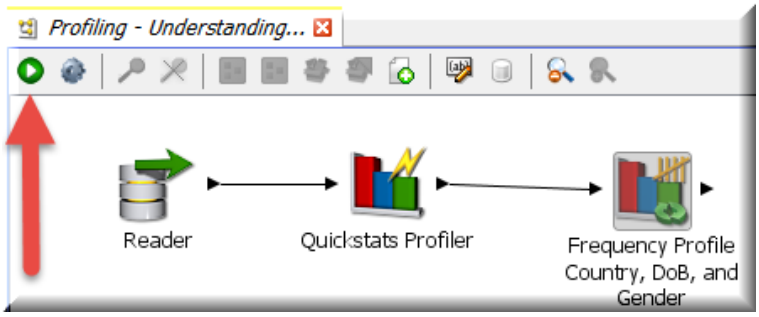
18. The **Frequency Profiler** dialog appears. Multi-select the **County**, **Town**, **DoB**, and **Gender** Available Attributes and select the icon to add the attributes to your **Selected Attributes**, then click **OK**



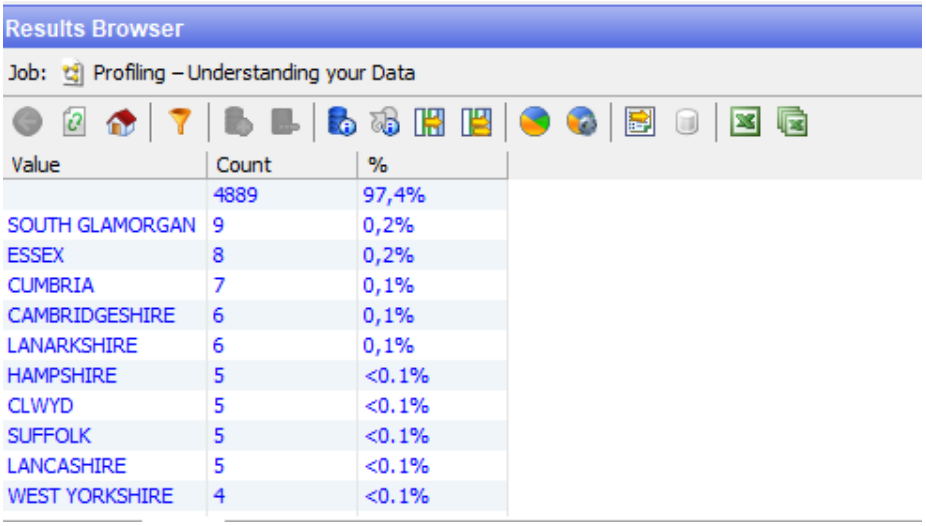
19. Processors can be renamed by double-clicking on the name of the processor within the canvas. Double click on the existing label of the **Frequency Profiler** and enter *Frequency Profile County, Town, DoB, and Gender* to rename the processor

    ** *Right clicking on the processor icon and selecting Rename also allows renaming of the processor*

20. Click the **Run** icon  to start the process as the Frequency profiler has yet to run and we want to view the results. Wait for execution to complete



21. Click the **Frequency Profile County, Town, DoB, and Gender** processor to view the results in the **Results Browser**. Notice the 4 distinct tabs at the bottom left corner: **County, Town, DoB, Gender and Data**. Let's take a moment to analyze what these different tabs tell us about our data set



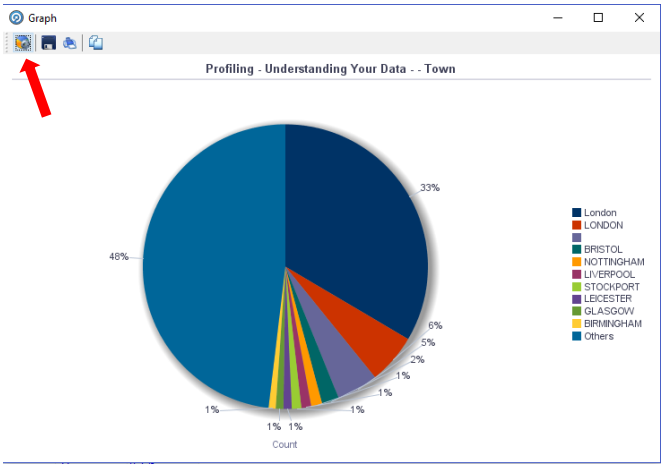
Value	Count	%
	4889	97,4%
SOUTH GLAMORGAN	9	0,2%
ESSEX	8	0,2%
CUMBRIA	7	0,1%
CAMBRIDGESHIRE	6	0,1%
LANARKSHIRE	6	0,1%
HAMPSHIRE	5	<0.1%
CLWYD	5	<0.1%
SUFFOLK	5	<0.1%
LANCASHIRE	5	<0.1%
WEST YORKSHIRE	4	<0.1%

*** Having separate tabs for each **Selected Attribute** is one of the many user interface features of EDQ – usability being one of the top 3 differentiators for EDQ*

This specific processor happens to tell us a lot about our data set just by observing the different values. Notice there are different representations for the same country name.

Finally, there are over 97% of the rows with no data in the County column. Later in the lab, we are going to populate empty County fields with an external Reference Data matching Post Code and County data.

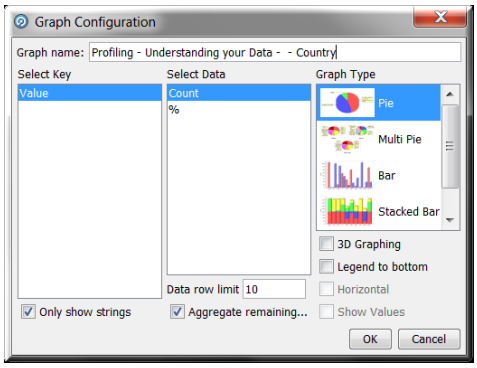
22. Navigate to Town data and Click the **Graph Results** graphical button  in the **Results Browser** to see a chart of the different values presented in the **Graph** dialog

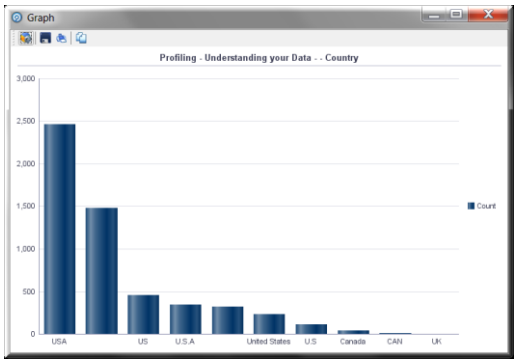


23. The first button in the toolbar as shown in the screenshot above, will allow you to change the title of the Chart, configure the Chart to a different type of visualization or

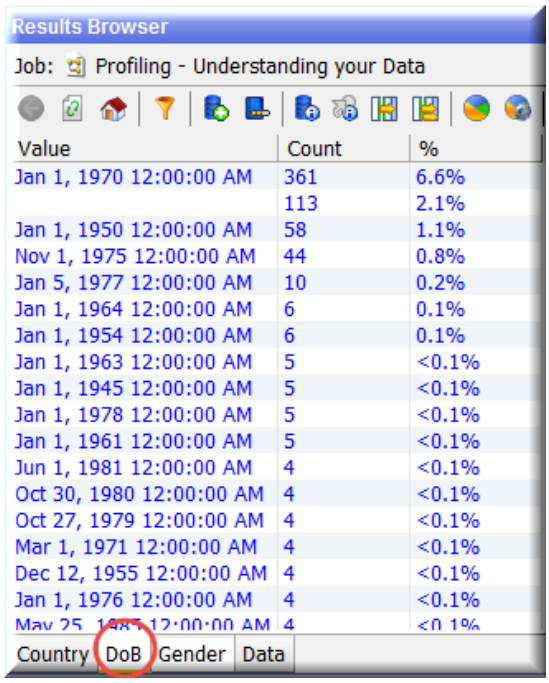
modify the type of data displayed. Click this button to open the Graph Configuration dialog.

24. Select the **Bar Graph Type**, then click **OK**

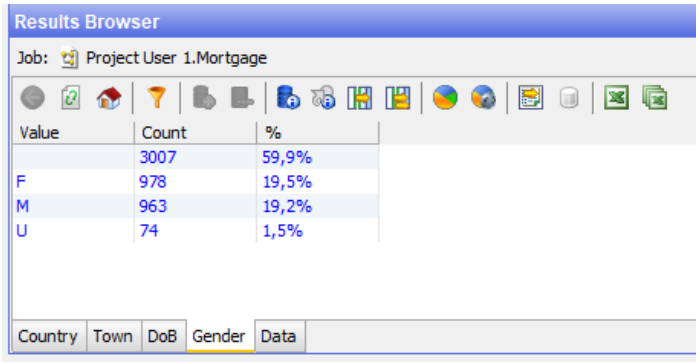




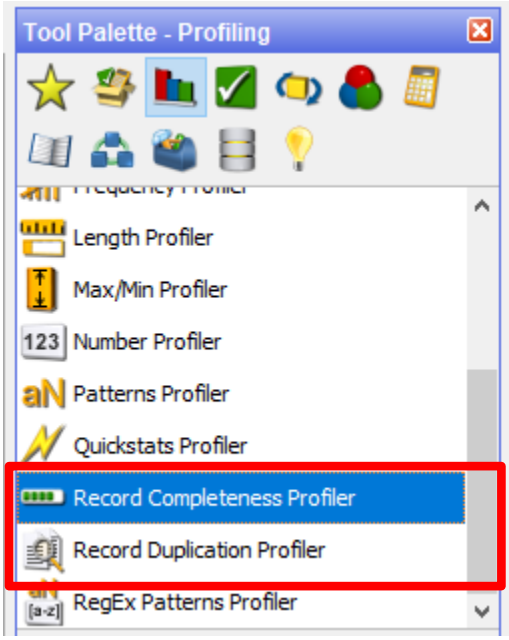
25. Close the **Graph** dialog window and click the **DoB** tab in the bottom of the **Results Browser** to view the results of the **Frequency Profiler**



26. Click the **Gender** tab in the bottom of the **Results Browser** to view the results of the **Frequency Profiler**



27. Let's add a few more processors to **Profiling – Understanding your Data** process. Return to the **Tool Palette** - **Profiling** and find the **Record Completeness Profiler** and **Record Duplication Profiler** by using the Search box. Type Record in the Search textfield to find the processors and drag and drop these processors to the Project Canvas.



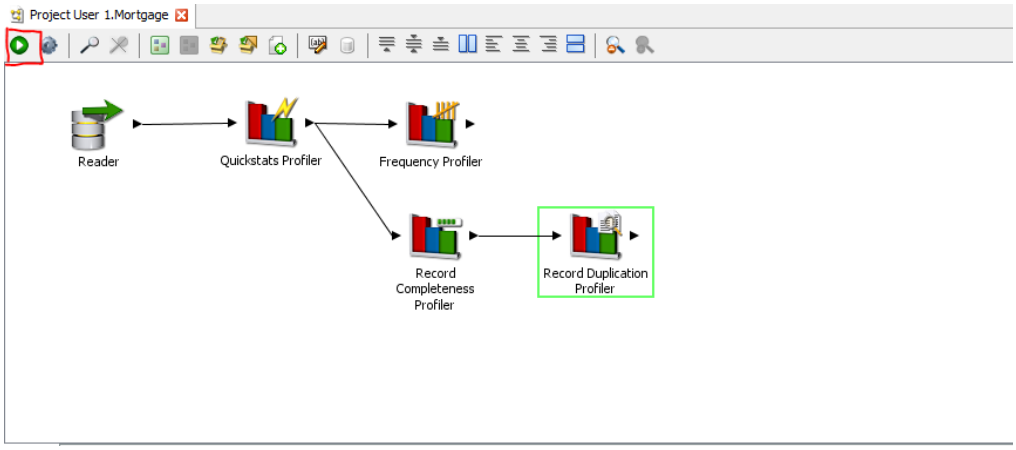
28. Click and drag from the output triangle of the **Quickstats Profiler** processor so that the connector line reaches the input triangle of the **Record Completeness Profiler**

29. The **Record Completeness Profiler** configuration dialog appears. Click the select all icon  to have all the data columns participate, then click OK

30. Click and drag the output triangle of the **Record Completeness Profiler** to the input triangle of the **Record Duplication Profiler**

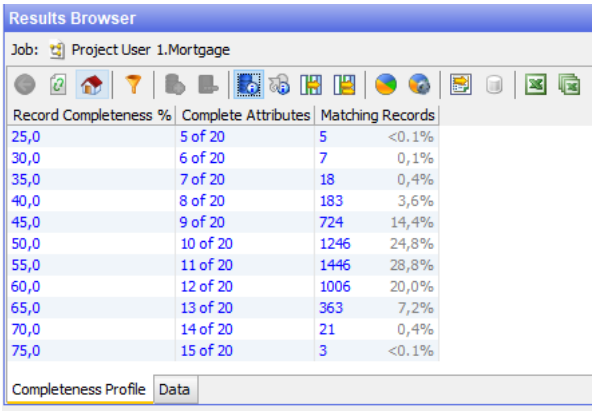
31. The **Record Duplication Profiler** configuration dialog appears. Click and select the **FullName** attribute from **Available Attributes** and click the icon  to move the ttribute to the **Selected Attributes**. Similarly, click and select the **Postcode** attribute, then click OK

32. Click the Run icon  in the toolbar to run the process



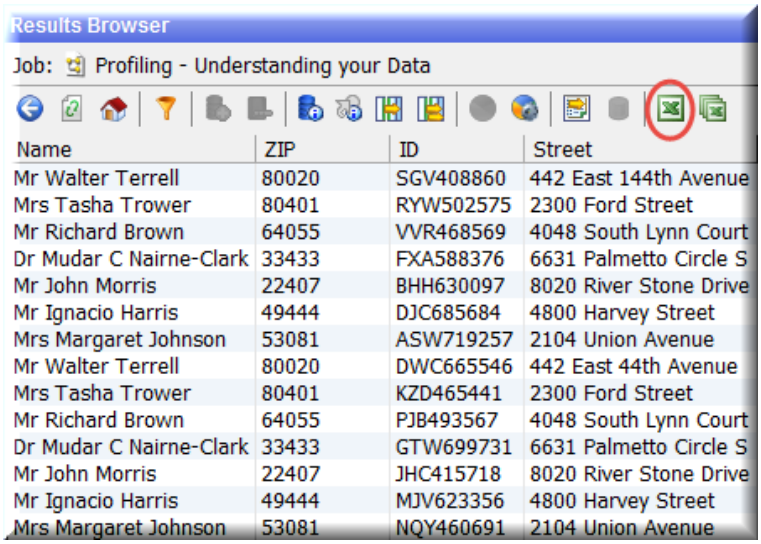
Just as the **Quickstats Profiler** provided many details about the dataset, the **Record Completeness Profiler** will analyze records with all of the selected attributes to display completeness. The **Record Duplication Profiler** will analyze records for duplicates across the selected Name and Postcode attributes.

33. Click on the **Record Completeness Profiler** processor to view the results in the **Results Browser**. You can see that only **1143** of the customers have all **10 of the 20** attributes filled. Click the **Show Additional Information** icon . Notice that those **1143** complete records only make up **22.8%** of the entire dataset.



34. Click on the **Record Duplication Profiler** to view the results. Drill down on the **339** representing Duplicated records. We learned how to raise issues earlier in the lab, you can also export the results to an Excel file to send to an individual in the organization for

further investigation. Click the **Export to Excel** icon in the **Results Browser** toolbar to save the file. (You do not need to save the file)



Name	ZIP	ID	Street
Mr Walter Terrell	80020	SGV408860	442 East 144th Avenue
Mrs Tasha Trower	80401	RYW502575	2300 Ford Street
Mr Richard Brown	64055	VVR468569	4048 South Lynn Court
Dr Mudar C Nairne-Clark	33433	FXA588376	6631 Palmetto Circle S
Mr John Morris	22407	BHH630097	8020 River Stone Drive
Mr Ignacio Harris	49444	DJC685684	4800 Harvey Street
Mrs Margaret Johnson	53081	ASW719257	2104 Union Avenue
Mr Walter Terrell	80020	DWC665546	442 East 44th Avenue
Mrs Tasha Trower	80401	KZD465441	2300 Ford Street
Mr Richard Brown	64055	PJB493567	4048 South Lynn Court
Dr Mudar C Nairne-Clark	33433	GTW699731	6631 Palmetto Circle S
Mr John Morris	22407	JHC415718	8020 River Stone Drive
Mr Ignacio Harris	49444	MJV623356	4800 Harvey Street
Mrs Margaret Johnson	53081	NQY460691	2104 Union Avenue

Lab 2: Auditing Your Data

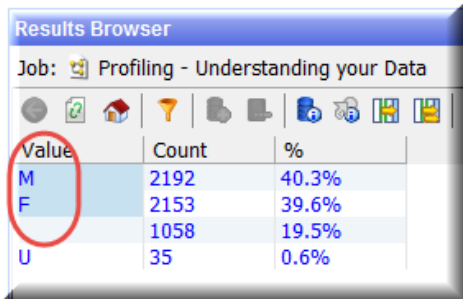
In this Lab, we will focus on EDQ Audit processors. Audit processors, or checks, check input data using business rules in order to assess whether or not it is fit for its business purpose.

Audit processors categorize each input record as to whether it is **valid** or **invalid** according to the **check.** Audit processors provide separate and easily accessible data output streams for valid and invalid records enabling separate workflows for handling valid and invalid records within an EDQ Process. Audit processors implicitly use the business rules that you apply to a given data attribute when profiling. For each type of business rule that you can apply, there is an Audit processor.

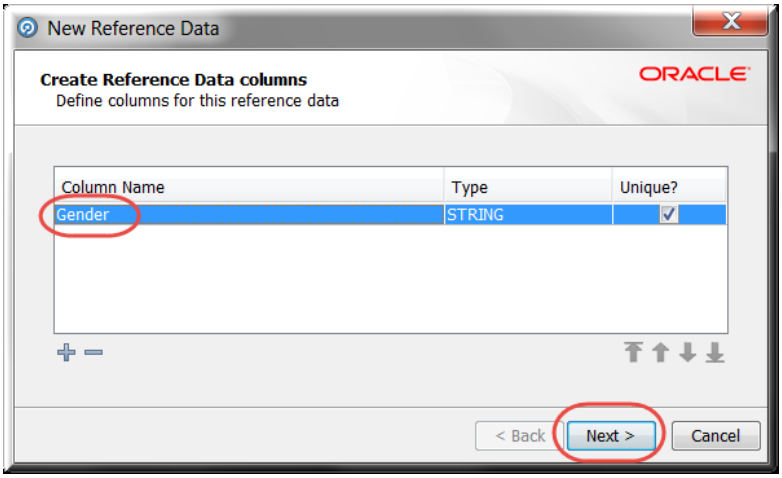
Create Reference Data

1. First, return to your **Profiling – Understanding your Data** process in the left-hand side **of Director**. Click on the *Frequency Profile Country, Town, DoB, and Gender*. Then click on the **Gender** tab in the bottom left corner of the **Results Browser**

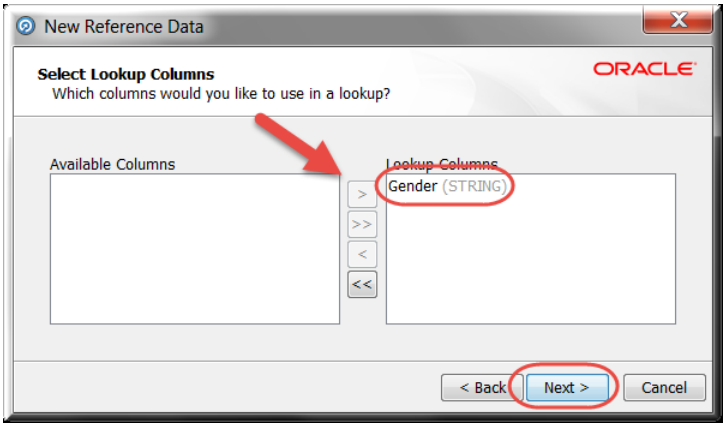
2. Hold down CTRL key and click on the **M** and **F** values

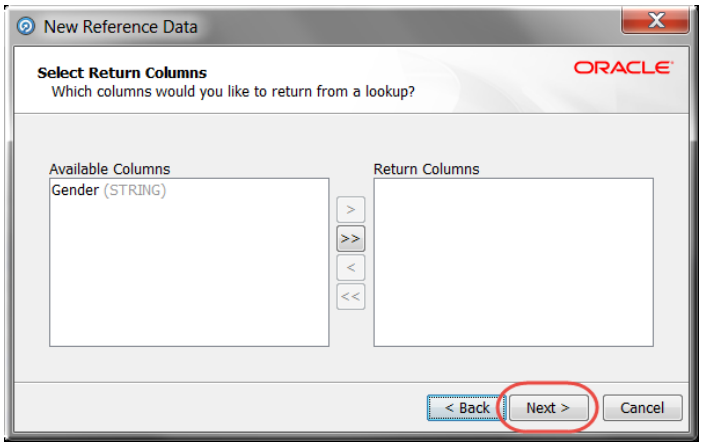


3. Right-click and select **Create Reference Data**. The **New Reference Data** dialog appears. Rename the attribute name to *Gender*, click **Next** to continue

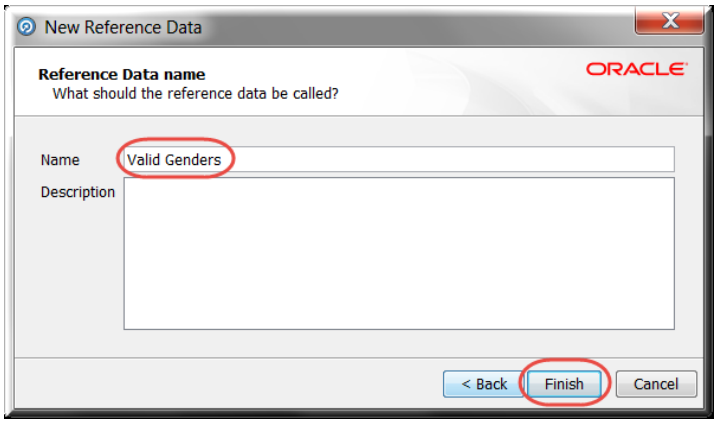


4. Add **Gender** to the **Lookup Column** using the > button, then click **Next** >. Click **Next >** on the next two screens to keep the default settings to continue (we will not add a return column or associate (classify) this reference data with any category)





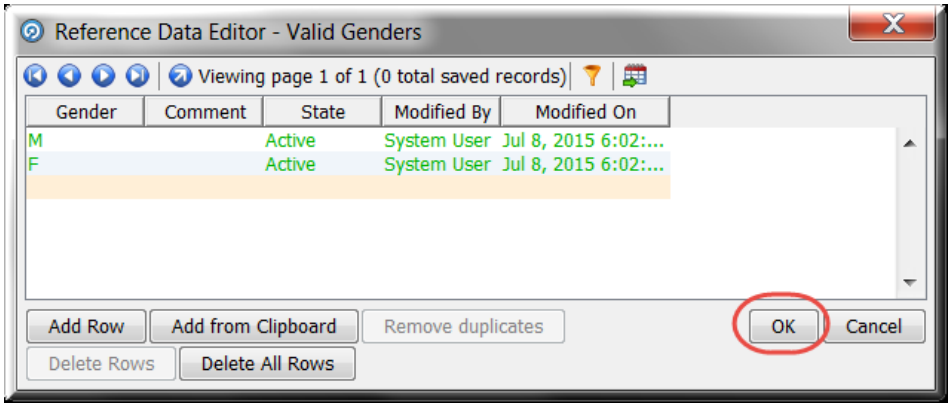
5. Finally, provide a name for this Reference Data: *Valid Genders*, then click **Finish**



6. The **Reference Data Editor** appears next

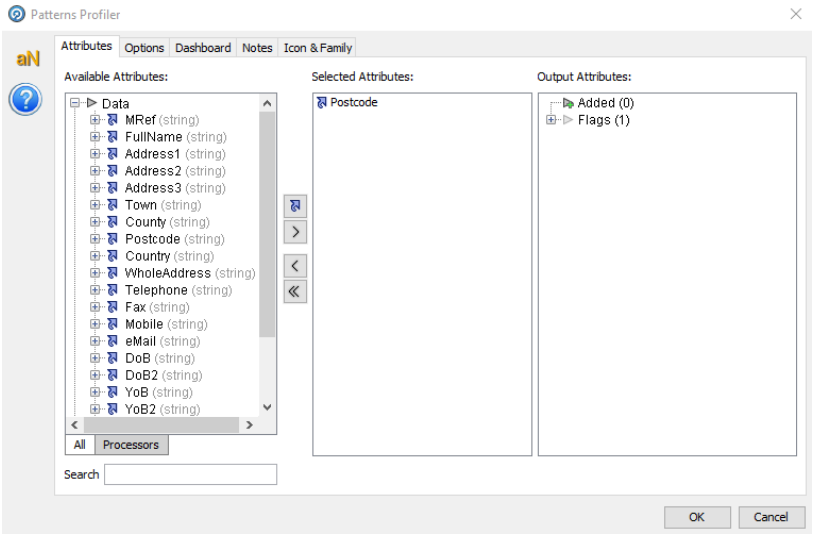
Here, you can modify the Reference Data to Add Rows or Delete Rows. EDQ comes with many different types of Reference Data out of the box which can dramatically speed up the time it takes to create data check processes. We will see some examples of the out of the box Reference Data in subsequent labs.

7. Here, you can see that EDQ has harvested the M and F values from your Profiling results to create your own custom Reference Data. We have just created the **single source of truth** definition of a **valid gender**. Click **OK** to return to the Project Canvas

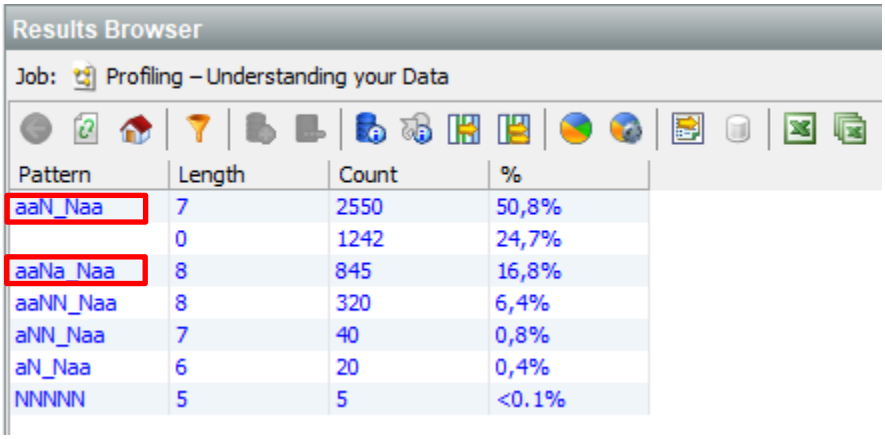


8. Navigate to the **Tool Palette** and find the **Pattern Profiler**. Drag and drop this into the **Project Canvas** and drag and drop the end triangle from the **Record Duplication Profiler** to the **Pattern Profiler**

9. The **Pattern Profiler** configuration dialog appears. Select **Postcode** from the **Available Attributes** and press the **>** button to add it to **Selected Attributes**, click **OK** to continue



10. Click the **Run** button  in the **Toolbar** in the top left corner above the **Process Tab**. After the Process completes, click the **Pattern Profiler** and view the results in the **Results Browser**

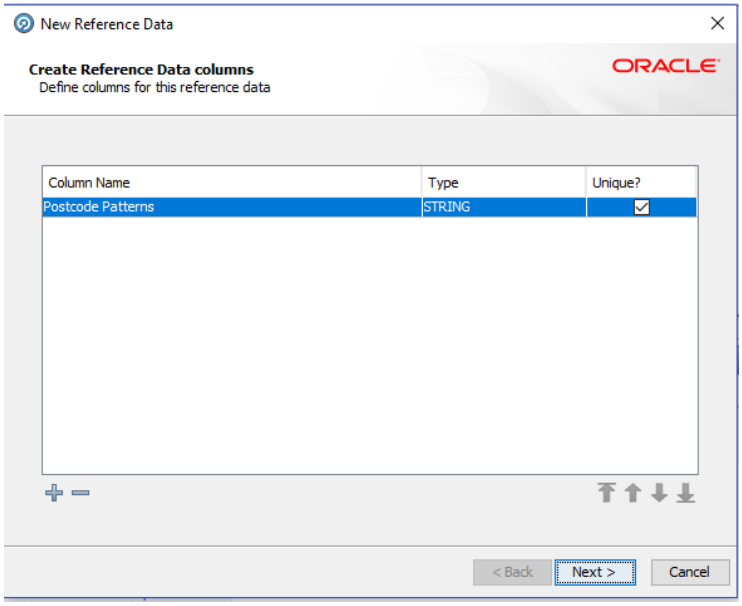


Pattern	Length	Count	%
aaN_Naa	7	2550	50,8%
	0	1242	24,7%
aaNa_Naa	8	845	16,8%
aaNN_Naa	8	320	6,4%
aNN_Naa	7	40	0,8%
aN_Naa	6	20	0,4%
NNNNN	5	5	<0.1%

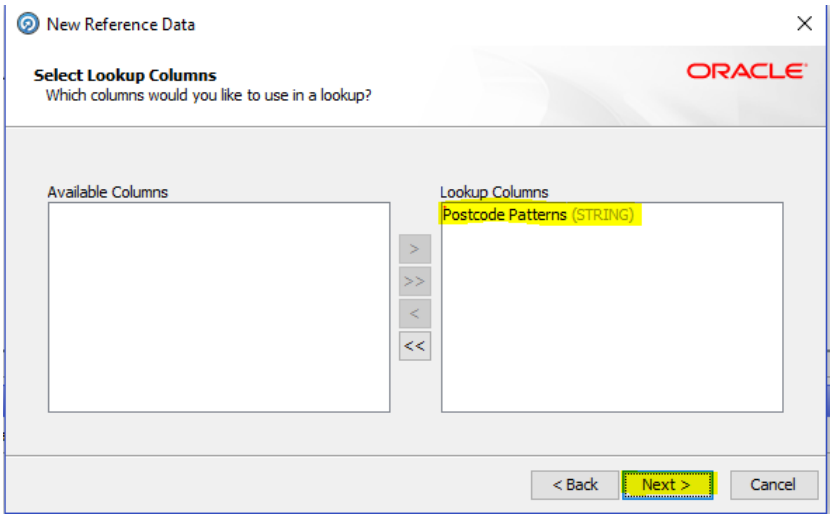
**** N signifies a number, p signifies punctuation, a signifies an alpha character, and _ signifies a space.*

11. Now that we have Patterns to create Reference Data from, repeat the steps taken when creating the Valid Genders Reference Data. Since we want 5 digit or 5 digit followed by 4 digits, CTRL click on **aaN_Naa** and **aaNa_Naa**

12. Right-click on the **aaN_Naa** and **aaNa_Naa** select **Create Reference Data**, and the **New Reference Data** Dialog appears. Rename the **Column Name** *to Postcode Patterns* by double clicking on **Pattern** under Column Name, click **Next >** to continue



13. Add **Postcode Pattern** to the **Lookup Columns** using the > button, then click **Next** > on the next two screens (we will not add return columns or select a category) to continue



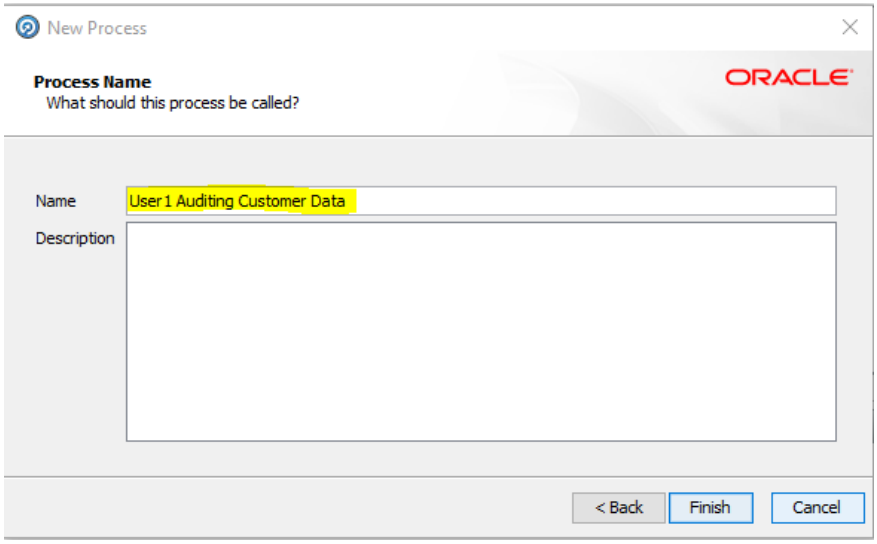
14. Last, give this New Reference Data a name ***Valid Postcode Patterns***, click **Finish**. The **Reference Data Editor** appears. If any values were selected by accident, remove those rows. Otherwise, click **OK** to continue. We have just created a **single source of truth** definition for the pattern of a valid Postcode.

Create Audit Process

We will now begin to create a new Process for Auditing our US Customer Data. The Reference Data we just created in the past few steps will be utilized by some of the out of the box Audit Processors within our Audit (data checking) Process.

15. Return to the **Project Browser** in the left side of your Director window, and underneath your **Project (Project User3)**, right-click on **Processes** and click **New Process**…

16. Select the **Staged Data User3**, then click **Next** >. Click **Next** on the next screen (we will not add any more profiling here). Name this new process ***User3 Auditing Customer Data***, then click **Finish** to continue

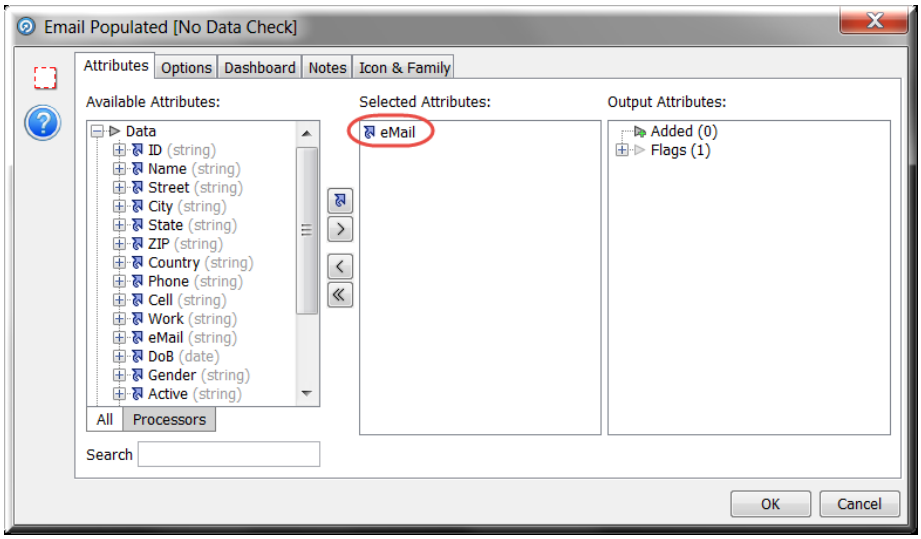


17. As with the first process we created, **a Reader Processor** is automatically added to the Project Canvas. Navigate to the **Tool Palette** to find the Audit icon.

    Take a moment to review the different Audit Processors. Hover your mouse-tip over the different entries to get a brief description.

18. First, drag and drop a **No Data Check** processor onto the Process canvas. Right click on the **No Data Check** processor and select **Rename** to re-name it to *Email Populated* and press the enter key. Drag and Drop the end triangle from the **Reader** to your newly named **Email Populated** Audit process.

19. The **Email Populated** configuration dialog appears. Select **eMail** from **Available Attributes** and add it to **Selected Attributes**, click OK to continue.



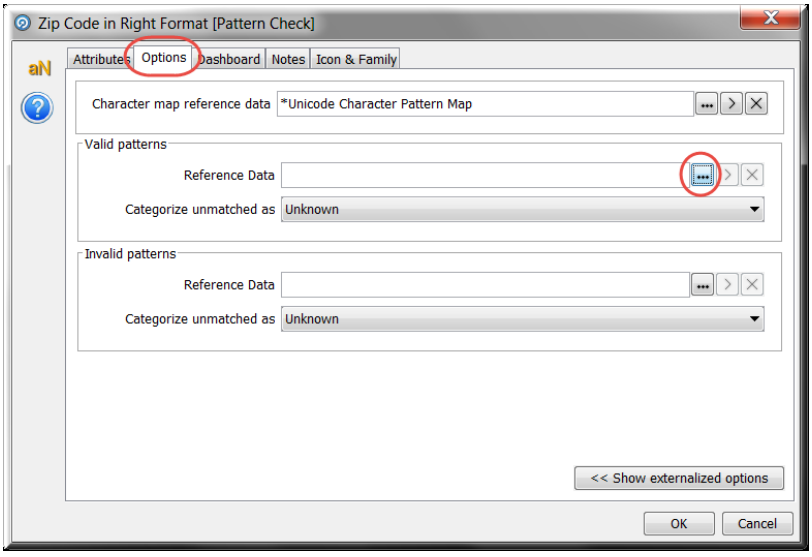
20. Click **Run** in the **Toolbar** above **User3 Auditing Customer Data** process tab and select the **Email Populated** audit processor to view the results.

    **** Note the **Without Data** and **With Data** values in the **Results Browser**. If desired, we can continue to develop this process using one or more of the end point output data stream triangles from the Processor by choosing **Data**, **No Data** or **All**.*

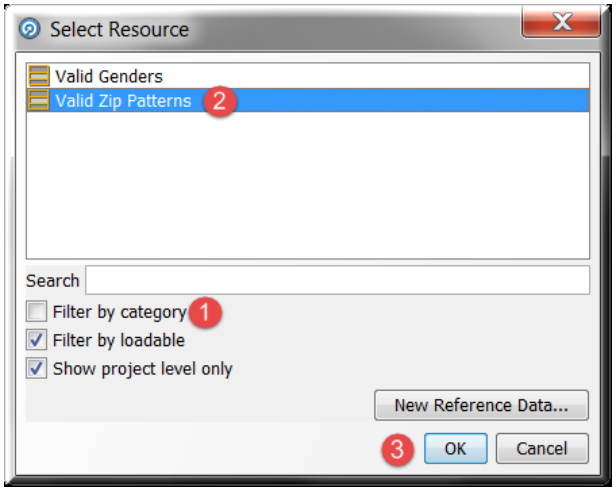
21. Next, find the **Pattern Check** Processor in the Tool Palette. Drag and drop it into the canvas and rename it to ***Postcode in Right Format*** by right clicking on the Pattern Check Processor and pressing the enter key

22. Connect **the All** end triangle from **Email Populated** to the **Postcode in Right Format** processor. The configuration dialog for the **Pattern Check (Postcode in Right Format)** processor appears. Select the **Postcode** from **Available Attributes** as the Field for validation.

23. Click the **Options** tab at the top of the dialog box, and then click the  button in the **Valid Patterns** section in the middle of the window



24. Uncheck **Filter by Category** in the **Select Resource** Window. This is where you will select the Reference Data we created for the different types of valid Postcodes (single source of truth for a valid Postcode format) at the beginning of the lab. Click on **Valid Postcode Patterns**, then click **OK.**

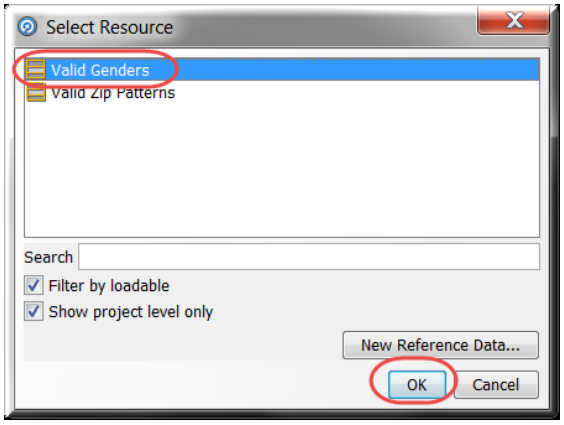


25. In the section under **Valid Patterns**, click the drop-down box to change **Categorize unmatched as** to **Invalid**, then click **OK** to continue
26. Click the **Run**  icon in the toolbar to start the process. Click the **Postcode in Right Format** to view the results.

**** Notice that there are 3.395 Valid Records and 1.627 Invalid Records. That is, there are 1.627 records that fail the 'fit for use' rule that do not match aaN_Naa or aaNa_Naa.*

27. Return to the **Tool Palette** and find the **List Check** processor. Drag and drop it onto the Project Canvas and link the **All** triangle from **Postcode in Right Format** to the **List Check** Processor.

28. Select the **Gender** attribute and add to the **Field for validation**. Then click the **Options** tab in the top of the dialog box to add the Reference Data for Genders (single source of truth definition for a valid gender) created at the beginning of this lab. Click the icon to the right of the **Reference Data** text field in the section for **Valid Values** to browse for Reference Data and select the **Valid Genders** Reference Data, then click **OK** to continue.



29. Click **OK** to close the **List Check** dialog box. Double-click the List Check processor to rename it to *Check for Valid Gender*. Finally, click the **Run** button in the toolbar to start the process.