Transform and Enrich Data in Oracle Analytics

Before You Begin

This lab shows you how to perform a set of manual and recommended data preparation actions to your dataset.

Background

Preparing and cleansing your data is an important step before visualizing a dataset. For example, the set might have sensitive data such as customers' social security numbers that you don't want to expose. You can hide or transform all characters of the social security number column, remove columns from a dataset, or extract portions of a data column to create a new column that contains the extracted data.

You can use the recommendations and available data preparation options in Oracle Analytics to improve data quality.

In this tutorial, you use a spreadsheet as the data source. You can add spreadsheet files that have the XLSX extension and that are no larger than 100 MB. You can also use comma-separated value (CSV) and text (TXT) files to create datasets. You can perform data preparation actions on supported data sources.
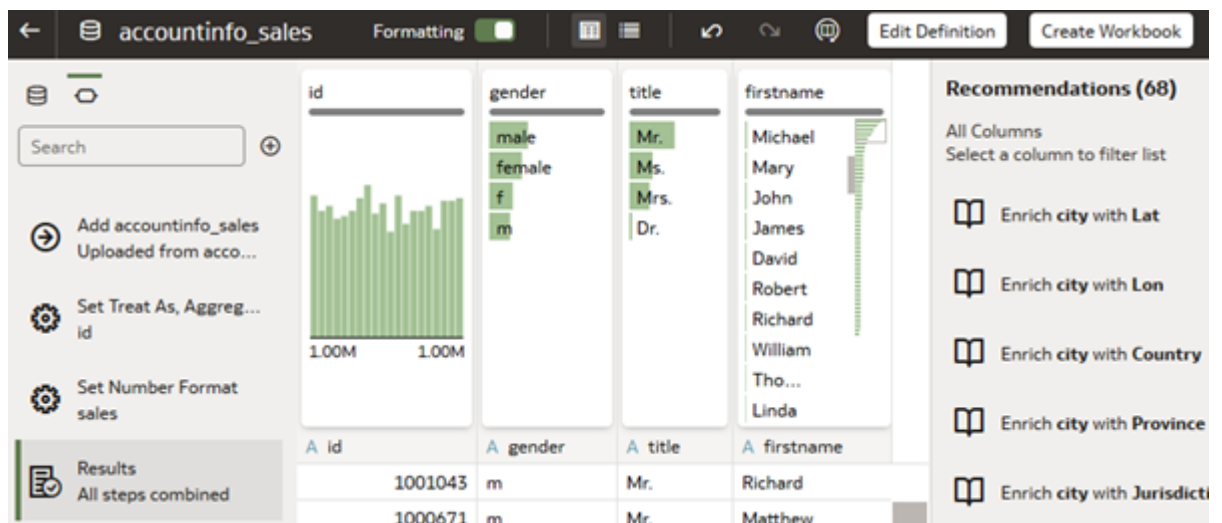
What Do You Need?

- Access to Oracle Analytics Cloud or Oracle Analytics Desktop
- Download the accountinfo_sales.xlsx file to your computer.

Create a Data Source

Oracle Analytics displays recommendations for the data, by column, in the dataset. In this tutorial, you accept some of the recommendations that are relevant for your analysis. You can also implement transformation changes for data in columns that don't have specific recommendations.

1. Sign in to Oracle Analytics.
2. On the **Home** page, click **Create**, and then click **Dataset**.
3. In **Create Dataset**, click **Drop data file here or click to browse**, select the accountinfo_sales.xlsx file, and then click **Open**.
4. In Create Dataset Table from accountinfo_sales.xlsx, click **OK**.
5. In the Join Diagram, click the **accountinfo_sales** tab.
6. In the Transform Editor, select the **id** column. In Properties, click **Measure** in the Treat As row, and then select **Attribute**.
7. Select the **Sales** column. In Properties, click **Number Format** ⊞. In the Number Format row, click **Number**, and then select **Currency**.

8. Click **Save**. In Save Dataset as, enter accounting_salesinfo, and then click **OK**.

Extract Data from a Column

When you extract data from a column, a new column is created that contains the extracted data. In this section, you extract the area code from phone numbers that use the North American Numbering Plan.

1. In the accountinfo_sales dataset, click **Toggle Quality Tiles** ☐ to close the insights over each column.
2. Select the **phone** column. In the **Recommendation** list, click **Extract area code from phone**.

   Oracle Analytics adds an area code column to the dataset.

Conceal Sensitive Customer Data

To comply with security policies for sensitive data, you can obfuscate all or a portion of the data in a column. If some users need to see the sensitive data, you can create a duplicate dataset containing the sensitive data.

1. Select the **ccnumber** column.
2. In the **Recommendations** list, click **Obfuscate First 12 Digits of ccnumber**.
3. Select the **ssn** column. In the **Recommendations** list, click **Obfuscate First 5 digits of ssn**.

Obfuscate First 5 Digits of ssn - From Recommendation •   Edit

A birthdate	A ccard	# ccnumber	# ccv	A ccexpires	A ssn
12/8/1970	MasterCard	############9058	222	6/2018	###-##-5630
6/14/1949	Visa	############5202	51	9/2015	###-##-5265
12/25/1983	MasterCard	############4645	354	3/2017	###-##-0733
4/15/1978	Visa	############9000	995	10/2014	###-##-6138
9/1/1939	Visa	############8937	770	4/2016	###-##-6380
7/30/1991	Visa	############0709	910	2/2016	###-##-5291
2/26/1974	MasterCard	############1863	548	1/2016	###-##-7941
5/10/1937	Visa	############4705	571	10/2014	###-##-4292
9/23/1937	Visa	############1007	525	9/2016	###-##-1071
4/26/1962	MasterCard	############7926	438	8/2014	###-##-1596
3/31/1931	MasterCard	############6281	786	5/2015	###-##-6765
5/25/1932	Visa	############7690	32	3/2015	###-##-7260
2/26/1980	Visa	############9396	698	3/2014	###-##-9920

Enrich Data with Geographic Coordinates

1. Select the **zip** column. In **Recommendations**, click **Enrich zip with Lat** (latitude).
2. In **Recommendations**, click **Enrich zip with Lon** (longitude).

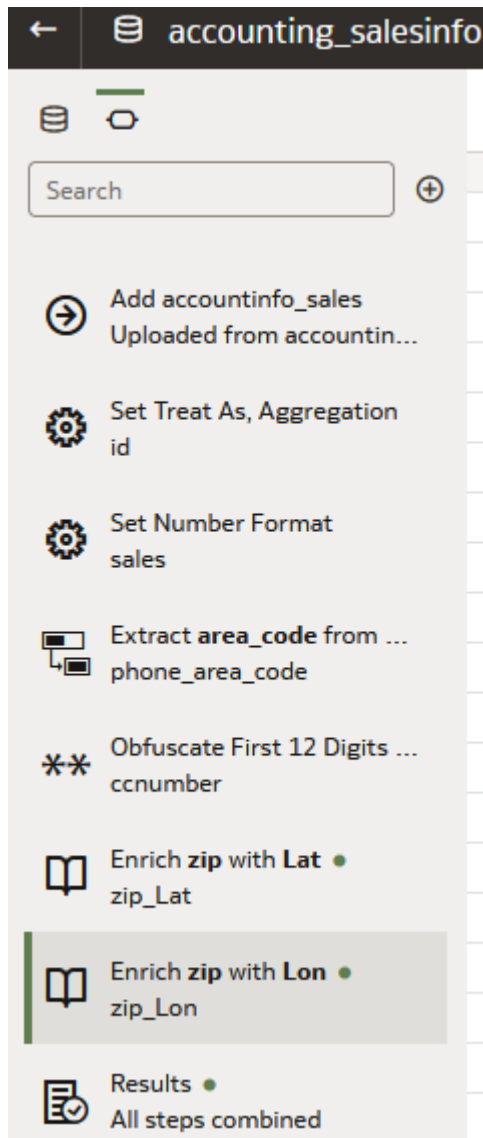Enrich zip with Lon - From Recommendation •   Edit   ← A zip (7)

A state	A zip	A zip_Lon	A zip_Lat
CO	80111	-104.8799	39.612299999999
NJ	07701	-74.0681	40.3584
NC	27597	-78.3174	35.832099999999
WA	99158	-117.2427	47.1282
NV	89121	-115.0902	36.123199999999
TX	75247	-96.8703	32.815199999999
CA	92614	-117.8298	33.682899999999
TX	77511	-95.251499999999	29.411999999999
TX	78401	-97.403	27.7941
FL	32216	-81.5831	30.2787
IL	60607	-87.657799999999	41.8721
VA	23505	-76.287499999999	36.9168
SD	57201	-97.123999999999	44.904299999999

Enrich zip with City

Enrich zip with State

Enrich zip with County

Enrich zip with Population_Largest_City

Enrich zip with Elevation_Meters

Enrich zip with Timezone

Enrich zip with State_Code

3. Click **Save** 🖫.

Your changes are listed in the Preparation Script ⟳ pane, and then applied when you save the dataset.

Inspect the Dataset

In this section, you review the changes implemented in the dataset.

1. Click **Go back** ⬅.
2. On the Home page, select the **accountinfo_sales** dataset, click the **Actions menu** ⋮, and then select **Inspect**.
3. In the dataset page, click **Data Elements** to view the columns added to the dataset.

accountinfo_sales
Dataset

<div style="text-align:right">Save</div>

General

Data Elements

Search

Access

Name	Data Type	Treat As	Aggregation
id	Number	Attribute	None
gender	Text	Attribute	None
title	Text	Attribute	None
firstname	Text	Attribute	None
middleinitial	Text	Attribute	None
lastname	Text	Attribute	None
street_address	Text	Attribute	None
city	Text	Attribute	None
state	Text	Attribute	None
zip	Text	Attribute	None
zip_Lon	Number	Attribute	None
zip_Lat	Number	Attribute	None
country	Text	Attribute	None
email	Text	Attribute	None