

## Data Profiling

- Overview
- Run analysis
- Histogram chart types
- Values
- Usage
- Suggestions

## Overview

Data profiling analyzes the data in your view, and determines what actions are available to optimize the efficiency of your view.



Ensure that you have the Data Profiling configuration setting enabled to use this functionality. Navigate to Administration > Admin Console > Data Sources > (click on your data source) > Usage Parameters > Data Profiling.

## Run analysis

A view can only be analyzed after the modeling stage. Make sure the view you're working with has already been **modeled** before attempting to run a data profiling analysis.

When profiling your data, you can run a basic profile by ignoring the advanced options, or you can set the advanced options according to your needs.

Options	Description
Rows to Profile	Choose how many rows of data to profile. More rows will give more relevant results, but will take longer to analyze.
Columns to Profile	Choose to run data profiling on all the fields in the view, or select a single field from the list.
Analysis Types	Select which data analysis you wish to run on the selected field(s) for suggestions to be tested and included.  Selecting none will still run the standard data profiling.

In the steps below, we will be working with a view that is being edited in the **Prepare** stage.

1. Ensure your view is in the **Prepare** stage

Yellowfin

Ski Team

Model Prepare

Agency

Athlete


Search

	Agency T...	Company...	Agency R...	Agency C...	Agency D...	Agency G...	Agency G...	Athlete ID	First Name	Last Name	Demogra...	Gender	Date of B...
1	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	9259	Alessandro	Fialo	Family	Male	01/01/1984
2	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	10210	Manuele	Pedroni	Family	Male	23/03/1990
3	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	12033	Kathryn	Easley	Family	Female	27/09/1989
4	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	9400	Gernaro	Cicciola	Family	Male	01/04/1993
5	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	12166	POZ	De Lorenzo	Family	Female	31/05/1971
6	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	11995	Mari	Yamada	Family	Female	26/08/1982
7	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	9284	Simone	Stollie	Family	Male	01/01/1984
8	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	12169	Martina	Florentino	Family	Female	31/08/1991
9	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	9644	Bjorn	Sieber	Family	Male	05/03/1976
10	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	10975	Julie	Robnett	Family	Female	01/01/1959
11	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	9191	Angelo	Connelli	Family	Male	01/01/1981
12	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	9339	Alberto	Pedrazzini	Family	Male	01/01/1969
13	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	12098	Regina	Zankl	Family	Female	17/07/1991
14	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	11792	Jessica	Neuner	Family	Female	21/03/1990
15	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	10986	Nicole	Cameron	Family	Female	01/01/1973
16	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	9909	Werner	Gegger	Family	Male	30/04/1972
17	Online	Planet Travel	Europe	Germany	Family	<GIS MultiPolyg...	DE	11098	Nata	De Leeuw	Relaxation	Female	02/02/1978

☐ Varchar
 ☐ Varchar
 ☐ Varchar
 ☐ Varchar
 ☐ Varchar
 ☐ GIS Geometry
 ☐ Varchar
 ☐ Integer
 ☐ Varchar
 ☐ Varchar
 ☐ Varchar
 ☐ Varchar
 ☐ Date

Raw Data ☐ Auto Refresh ☐

Rows 20

- Click on the **Data Profiling** tool  from the top toolbar to open the **Data Profiling** dialog box

Data Profiling

Data Profiling analyses the data in your view, and determines what actions are available to optimise the efficiency of your view. Examples include grouping fields, filtering out empty values, etc.

Run analysis over 1000 rows

Advanced ↓

Start Profiling

- Click on the **Advanced** link to expand the list of options

Data Profiling

Data Profiling analyses the data in your view, and determines what actions are available to optimise the efficiency of your view. Examples include grouping fields, filtering out empty values, etc.

Rows to Profile

Specify how many rows of data to profile. More rows will give more relevant results, but will take longer to analyse.

1000

Columns to Profile

Run profiling on all columns in the view.

All

Select

Analysis Types

Select which types of analysis to run on the view data.

All

Date Grouping

GeoPack

Start Profiling

- Use the table above to change any of the three main advanced options to suit your needs
- Click on the **Start Profiling** button to commence the analysis

Once the analysis is complete, your view will display histograms under each field

Model

Prepare

Test

More

Publish

Agency T...Company...Agency R...Agency C...Agency D...Agency G...Agency G...

Athlete IDFirst NameLast NameDemogra...GenderDate of Bi...

OnlinePlanet TravelEuropeGermanyFamily<GIS MultiPolygo...DE9259AlessandroFalloFamilyMale01/01/1984

OnlinePlanet TravelEuropeGermanyFamily<GIS MultiPolygo...DE10210ManuelePedroniFamilyMale23/03/1990

OnlinePlanet TravelEuropeGermanyFamily<GIS MultiPolygo...DE12033KathrynEastleyFamilyFemale27/09/1989

OnlinePlanet TravelEuropeGermanyFamily<GIS MultiPolygo...DE9490GennaroCiotolaFamilyMale01/04/1993

OnlinePlanet TravelEuropeGermanyFamily<GIS MultiPolygo...DE12166POZDe LorenzoFamilyFemale31/05/1971

OnlinePlanet TravelEuropeGermanyFamily<GIS MultiPolygo...DE11995MamiYamadaFamilyFemale26/08/1982

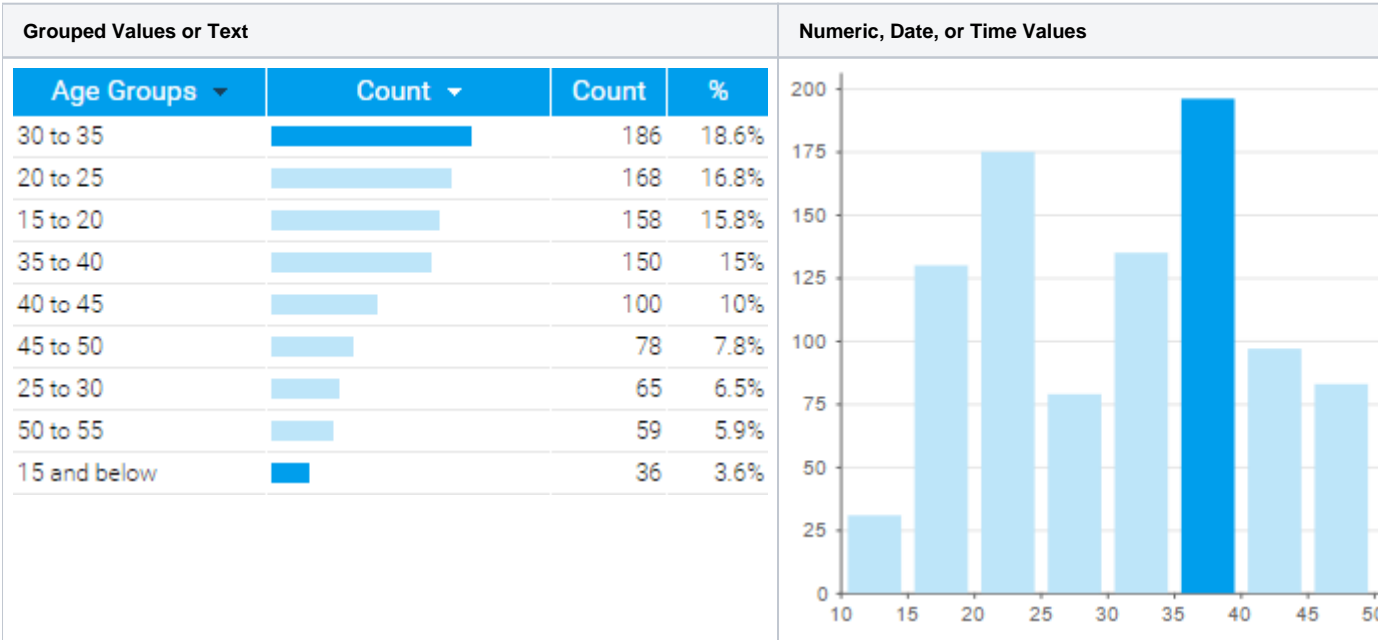
OnlinePlanet TravelEuropeGermanyFamily<GIS MultiPolygo...DE9284SimoneSibilleFamilyMale01/01/1984

See the sections below for further information on what information you can obtain from the histograms.

[top](#)

## Histogram chart types

A histogram chart analyzes the frequency of distinct values in the data and displays it at the top of each profiled field.



[top](#)

## Values

Sample Size:	1,000
Distinct Values:	682
Empty Values:	0
Min value:	9,026
Avg value:	10,404
Max value:	12,230
Median value:	10,111
Standard deviation:	990
Outliers:	0

Statistic	Description
Sample Size	Displays a count of the records within the field.
Distinct Values	Displays a count of the distinct values within the field.
Empty/NULL Values	Displays a count of the number of empty values within the field.
Numeric Fields	
Median	Displays the number separating the higher half of the sample from the lower half.

<b>Average</b>	Displays the mean value.
<b>Standard Deviation</b>	Displays the measure of the dispersion of a set of values.
<b>Minimum</b>	Displays the lowest value.
<b>Maximum</b>	Displays the highest value.
<b>Date/Time Fields</b>	
<b>Minimum</b>	Displays the earliest date.
<b>Maximum</b>	Displays the latest date.

[top](#)

## Usage

Reports ▲	Last Modified ▲	Usage ▲
Athlete Demographic Breakdown	29/3/2016	<div></div>
Athlete Figures by Location	29/3/2016	
Booking Process Donut	29/3/2016	<div></div>
Broken Bones	29/3/2016	<div></div>
Camp Region GIS Map	29/3/2016	<div></div>
Camp Revenue Detail Report	29/3/2016	
Camp Revenue Detail Report	29/3/2016	
Cancelled Bookings	29/3/2016	<div></div>
Customer Age & Location Breakdown	29/3/2016	<div></div>
Female Demographic	29/3/2016	<div></div>
Male Demographic	29/3/2016	<div></div>
Profitability by Customer Age & Location Breakdown	29/3/2016	<div></div>

Section	Description
<b>Reports</b>	Displays a list of reports that make use of the selected field.
<b>Last Modified</b>	Displays the last modified date of each listed report.
<b>Usage</b>	Displays the usage rate of each listed report.

[top](#)

## Suggestions

Depending on the type of field being profiled, Yellowfin may make suggestions as to functions that could be applied to it based on the outcomes of the analysis.

Suggested Actions (1)

Create Grouped Column

Create a new Grouped Value column based on this field.

Suggestion	Description
Reference Code Check	The values in the field will be analysed and compared to existing ref codes. It may then be suggested that a ref code be applied, updated, or created.
Null Check	The values in the field will be analysed and it may be suggested that null values be replaced or filtered out.
Number Uniqueness	The values in the field will be analysed and it may be suggested that the fields be grouped.
Date Uniqueness	The values in the field will be analysed and it may be suggested that the field be used as part of a date hierarchy.
Geography Check	The values in the field will be analysed and it may be suggested that the field be linked to a GeoPack.
Date Hierarchy	The values in the field will be analysed and it may be suggested that the field be linked in a drill down date hierarchy.