# Data Profiling

## Overview

top

### Run Analysis

1.  Click on the **Data Profiling** button

2.  If you want to profile all the fields in the view, click **Start Profiling**, otherwise click the **Advanced** link.
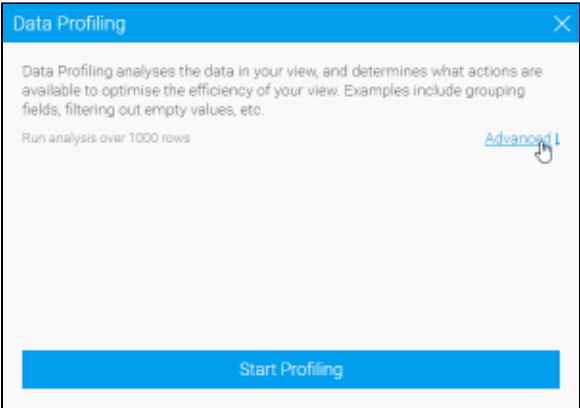
3.  Configure the advanced options and click **Start Profiling** to begin.

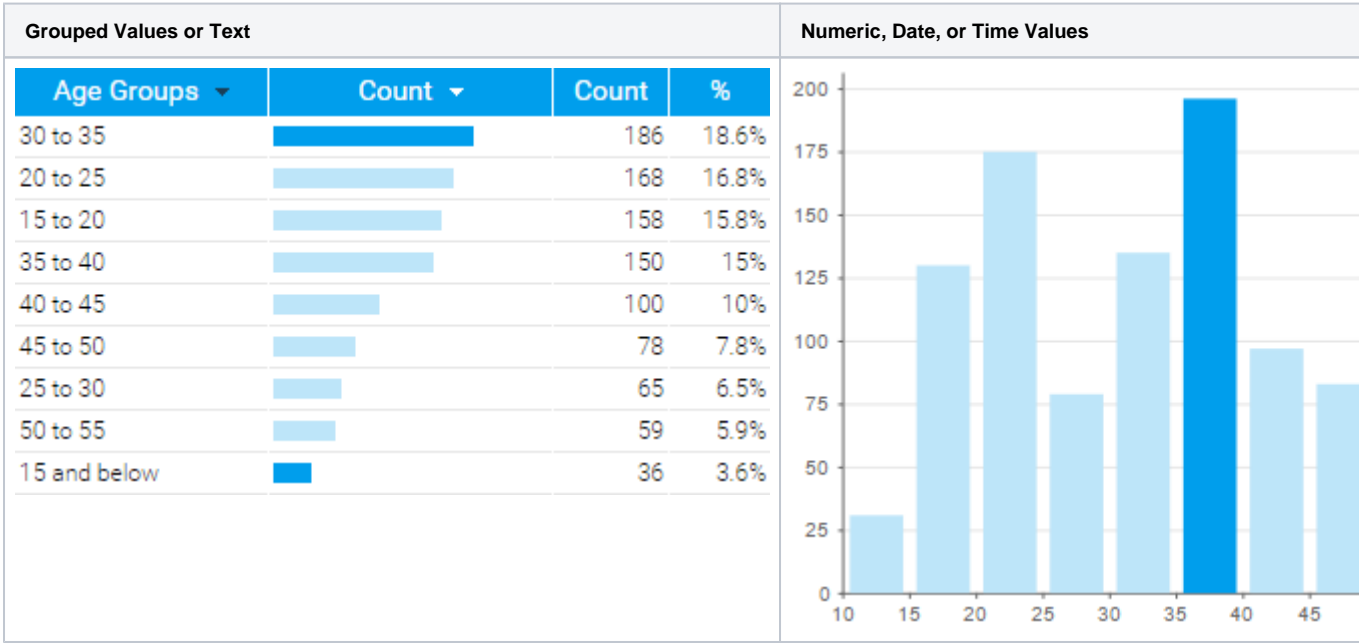| Options | Description |
|---|---|
| **Rows to Profile** | Specify how many rows of data to profile. More rows will give more relevant results, but will take longer to analyse. |
| **Columns to Profile** | Specify if you wish to run data profiling on all the fields in the view, or just select a single 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. |
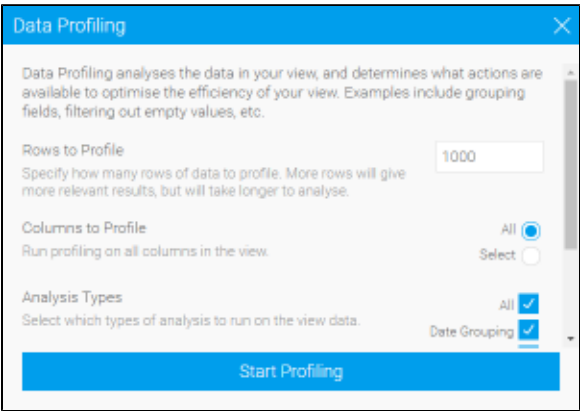
## Histogram

top

A histogram, analysing the frequency of distinct values in the data, is displayed at the top of each profiled field.

# Chart

| Grouped Values or Text | | | | Numeric, Date, or Time Values |
|---|---|---|---|---|

| Age Groups ▼ | Count ▼ | Count | % |
|---|---|---|---|
| 30 to 35 | | 186 | 18.6% |
| 20 to 25 | | 168 | 16.8% |
| 15 to 20 | | 158 | 15.8% |
| 35 to 40 | | 150 | 15% |
| 40 to 45 | | 100 | 10% |
| 45 to 50 | | 78 | 7.8% |
| 25 to 30 | | 65 | 6.5% |
| 50 to 55 | | 59 | 5.9% |
| 15 and below | | 36 | 3.6% |

# 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. |
|---|---|
| **Average** | Displays the mean value. |
| **Standard Deviation** | Displays the measure of the dispersion of a set of values. |
| **Minimum** | Displays the lowest value. |
| **Maximum** | Displays the highest value. |
| **Date/Time Fields** | |
| **Minimum** | Displays the earliest date. |
| **Maximum** | Displays the latest date. |

## Usage

top



| Section | Description |
|---|---|
| **Reports** | Displays a list of reports that make use of the selected field. |
| **Last Modified** | Displays the last modified date of each listed report. |
| **Usage** | Displays the usage rate of each listed report. |

## Suggestions

top
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. |