



compellon<sup>TM</sup>

Iris Pro



## **An Analytics Advisory Tool**

Streamline your data science process, drastically reduce time to modeling and quickly evaluate the information in your data to determine how well it supports your business analysis.

# Wasting time in your ML process?

- Identifying variables that matter can be one of the most time consuming processes in the ML process
- Training a model to only find out you missed the mark is timely and costly process

Reduce time to  
answers & quickly  
find what matters.

- Wade through 1000's of variables quickly to find the ones that are relevant to the target
- Understand the relationship of variables in regards to the target
- Speed time in training process in building stable accurate models by using variables that matter from the beginning

**Save time by cutting down your data processing and feature selection work to start building better models faster.**

Iris Pro answers the following questions:

Are the answers in  
your data?

What other questions  
can your data answer?

What set of variables are  
required to train my model?

How are those variables  
related to my objective?

# Are the answers in your data?

**When you need to know if your data supports your business objective.**

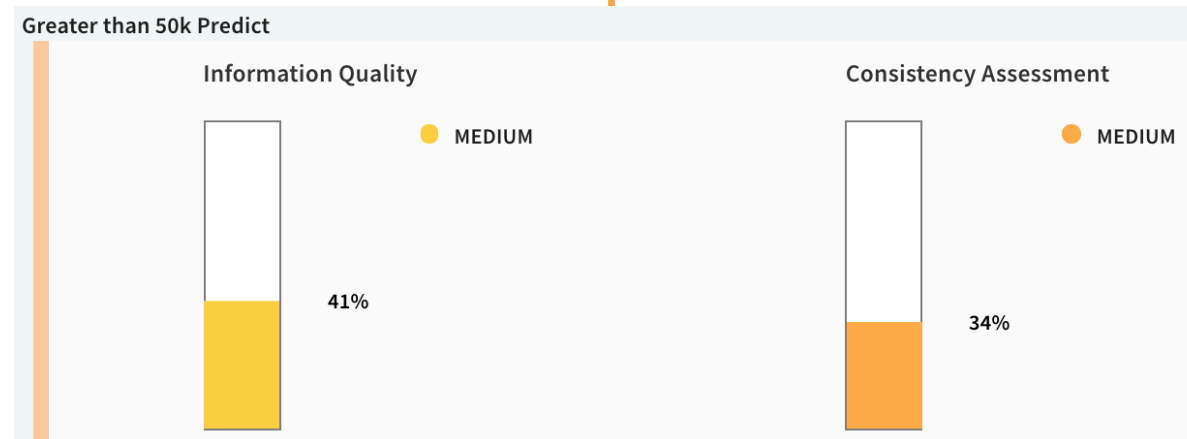
Using measures of information and consistency, Iris Pro evaluates your data's ability to build a model for your target.

## Information Quality

The amount of meaningful information in your data describing your target.

## Consistency

Is the information acting the same throughout, or are there hidden groups within the data.



# What variables are required to train your model?

Using methods from information theory we identify the maximum information in your dataset that explains your objective.

## Finds Best Group of Variables

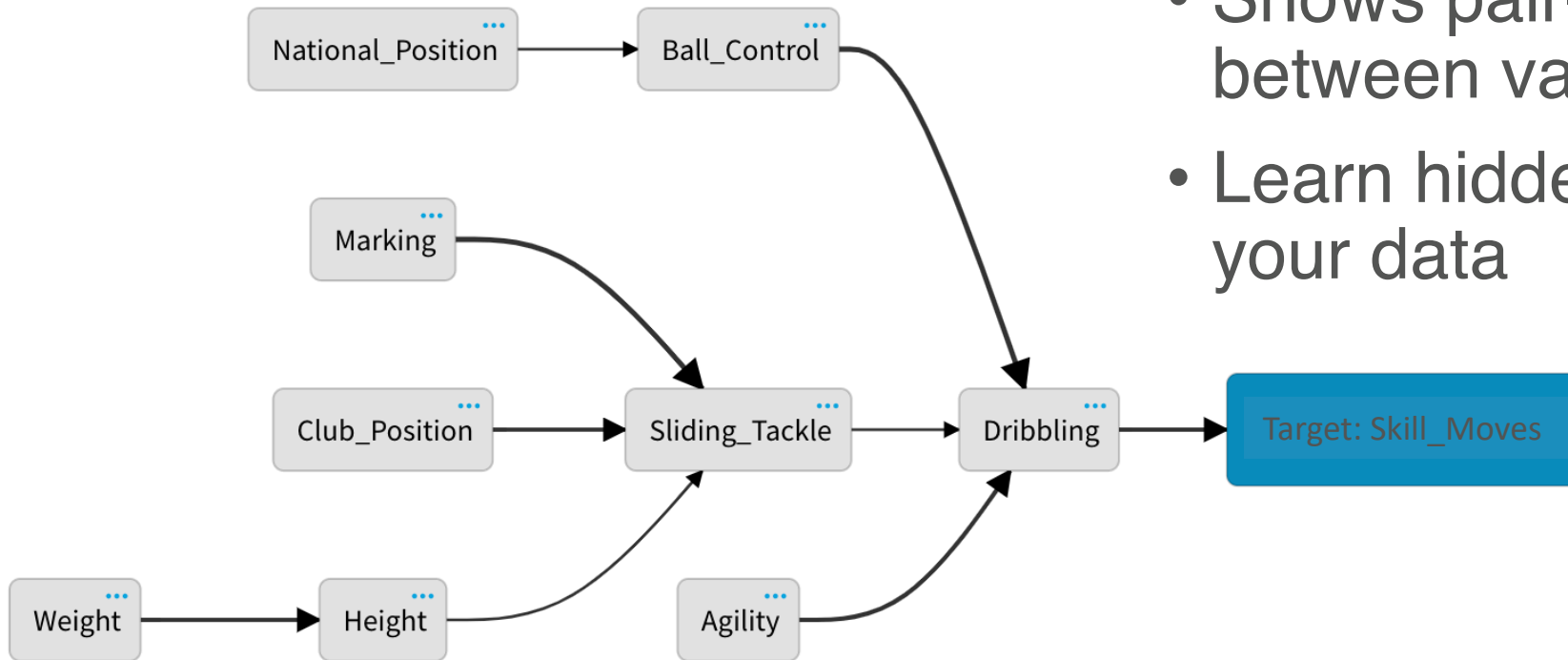
Removes overlapping information and noise while identifying unique contributors to your target outcome

## Benefits to Approach

- Independent of variable relationship
  - Scale Invariant
- Removes need to feature engineer
- Assumption Free Selection Approach
  - Model Approach Agnostic

# How are those variables related to my objective?

- Relationship of variables to the target objective
- Shows pair-wise relationship between variables
- Learn hidden relationship in your data





# What other questions can your data answer?

**When you have data and wonder what other business objectives it can solve.**

Using the same measures as before, Iris Pro identifies alternative target candidates and creates an informational graph of your data.

## Central Target

The center of the information graph showing how the data relates together. May or may not be valuable for analysis.

## Alternate Targets

Variables identified as good candidates to be targeted for analytical purposes.

## Information Graph

A graph representing the hierarchies of relations of variables in the dataset.

# Additional value from Iris Pro

**Iris Pro also identifies:**

Non-Informative Columns,  
Duplicate Columns, and  
Proxies to the Target.

**Know what you can exclude, and which columns have  
redundant information.**

# Iris Pro UI

# Dashboard

Datasets can be viewed by grid or list format.

To logout, click here.

New datasets can be added by clicking on the plus sign on the top left of the navigation bar.

You can return to dashboard anytime by clicking "Datasets."

Highlighted Dataset information shows on right side of dashboard.

The screenshot displays the Iris Pro dashboard interface. The top navigation bar includes the 'iris PRO' logo, an 'Add new' button (plus sign), a 'Delete' button (trash icon), and tabs for 'Review' and 'Analyze'. A search icon and a user profile icon are also present. The left sidebar features a 'Datasets' link. The main area shows a grid of dataset cards, each with a filename, row count, column count, and a 'NEW' badge. The 'churn.csv' dataset is highlighted. On the right, a detailed view of the 'churn.csv' dataset is shown, including its description, added date, row count, column count, and size. The bottom of the dashboard includes a 'Need Help?' link and the 'powered by compellon' logo.

Dataset Name	Rows	Columns
churn.csv	7043	21
Alteryxchurn.csv	7043	21
early_2012_2013_train-proc-reduced.csv	10000	30
loan-sampled10KRows.csv.zip	10000	74
soccerdemo.csv	17588	53
Yoga Studio Survey (1-Train).csv	1734	26
train.csv	41566	128
Kaggle_HRAnalytics.csv	1470	35
Airport Survey (1-Train).csv	5802	15
Auto Insurance Sales (All).csv	5802	15
UCI_Credit_Card.csv.zip	30000	25
Saas Customer Churn (1-Train).csv	1717	20
Combined Transaction Sheets 17-19 Items Out.csv	3477	61

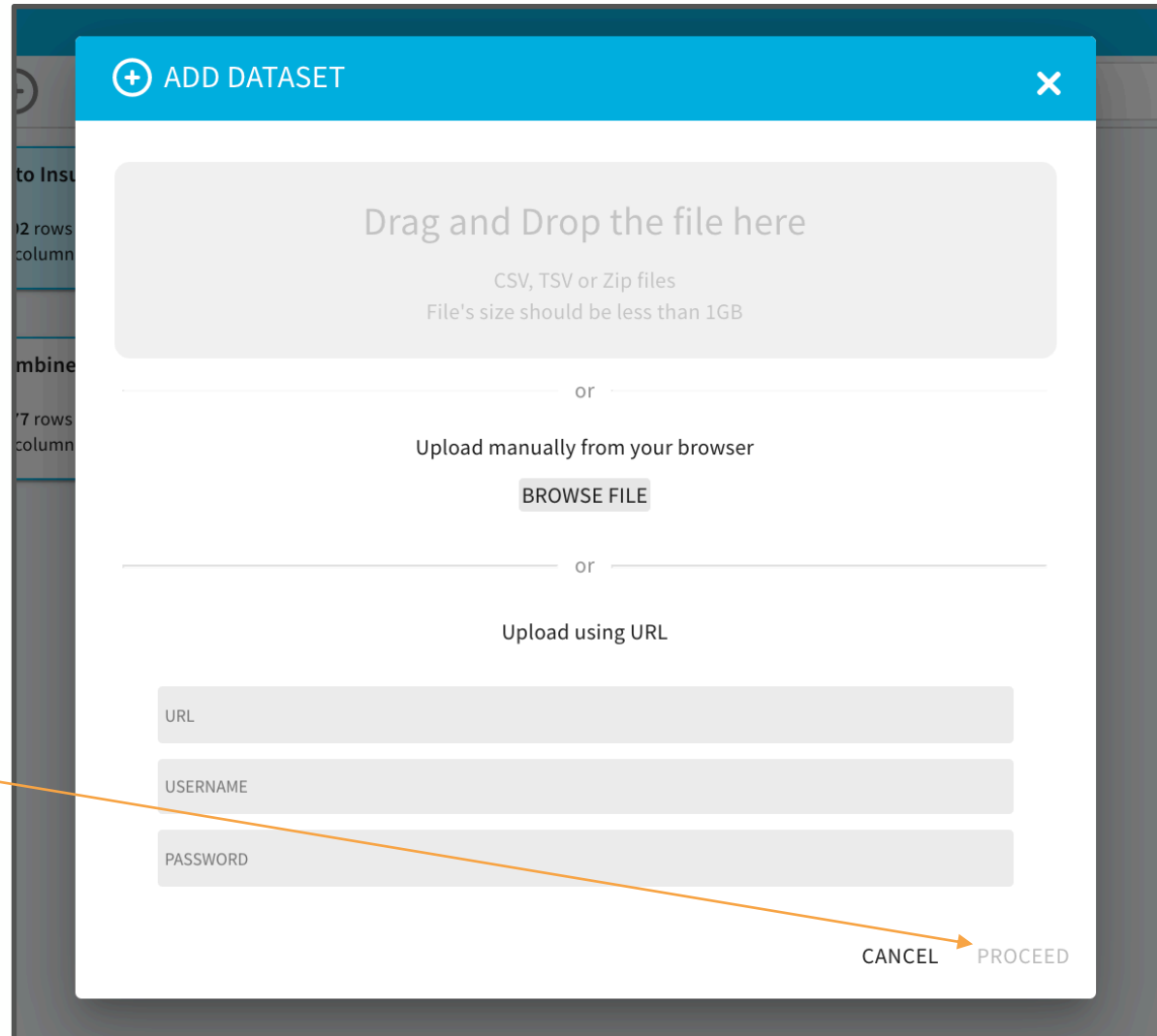
Property	Value
Added	11/12/2019
Rows	7043
Columns	21
Size	954.59 KB

# Dataset Upload

New datasets can be added using drag and drop, manual upload, or via URL.

We recommend Zip files for best upload.

Once your file is chosen, click on the “Proceed” button

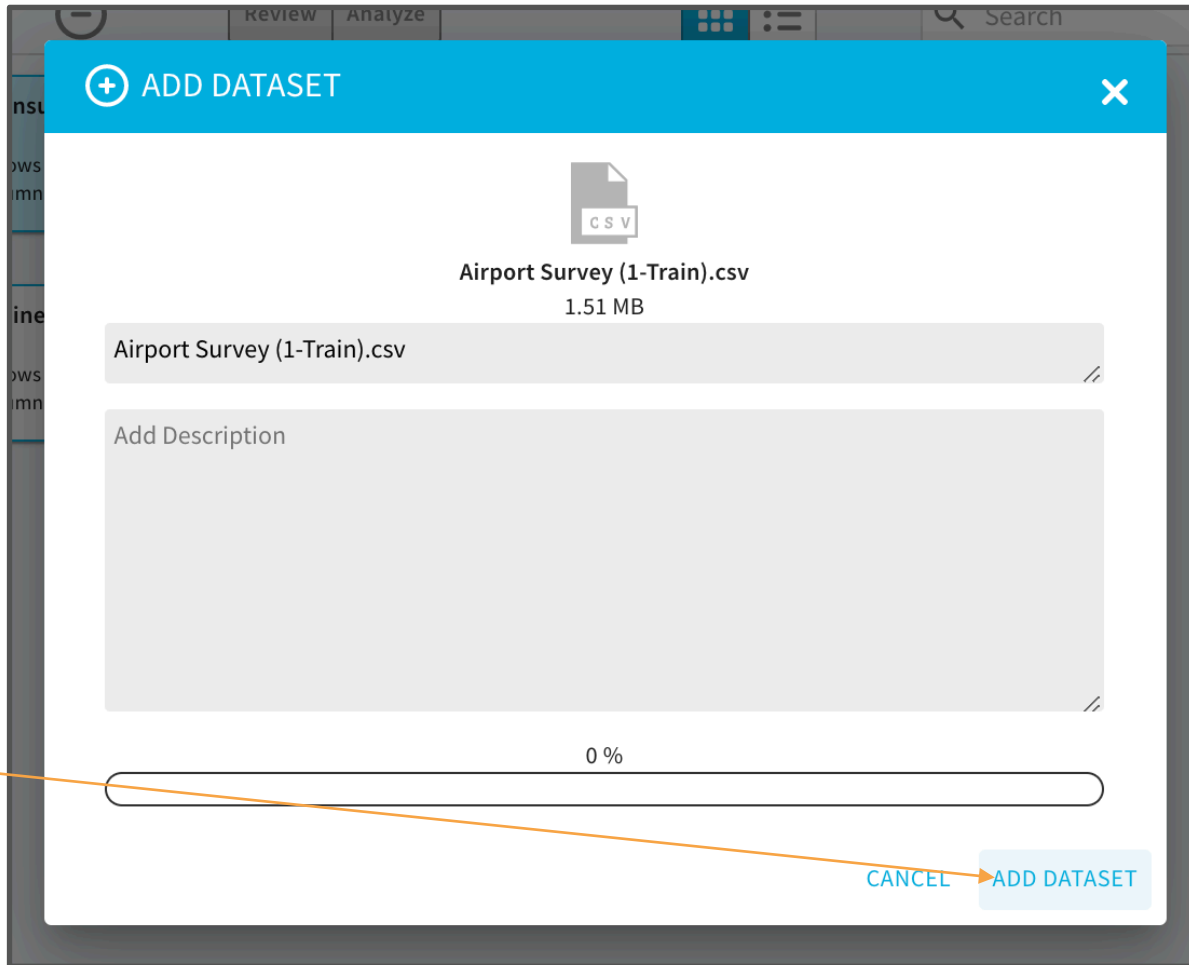


The screenshot shows a modal window titled "ADD DATASET" with a close button (X) in the top right corner. The main content area has a large grey box with the text "Drag and Drop the file here" and "CSV, TSV or Zip files" and "File's size should be less than 1GB". Below this is a horizontal line with "or" in the center. Underneath is the text "Upload manually from your browser" followed by a "BROWSE FILE" button. Another horizontal line with "or" follows. Below that is the text "Upload using URL". There are three input fields labeled "URL", "USERNAME", and "PASSWORD". At the bottom right, there are two buttons: "CANCEL" and "PROCEED". An orange arrow points from the text box on the left to the "PROCEED" button.

# Dataset Upload

The second modal is where you can rename your file and add a description. You also have this capability on the dashboard.

Once you have made any edits, click on the “Add Dataset” button to trigger the upload.



The screenshot shows a modal window titled "ADD DATASET" with a close button (X) in the top right corner. Inside the modal, a file icon labeled "CSV" is displayed above the text "Airport Survey (1-Train).csv" and "1.51 MB". Below this, there is a text input field containing "Airport Survey (1-Train).csv" and a larger text area labeled "Add Description". At the bottom of the modal, there is a progress bar showing "0 %". To the right of the progress bar are two buttons: "CANCEL" and "ADD DATASET". An orange arrow points from the text box on the left to the "ADD DATASET" button.

# Reviewing Datasets

Select the dataset you wish to review, it will highlight in blue.

Click on “Review” button in navigation bar to move to review page.

The screenshot displays the Iris Pro interface for managing datasets. A grid of dataset cards is shown, each with a file icon, name, row count, and column count. The 'churn.csv' dataset is highlighted in blue. The navigation bar at the top includes a 'Review' button (eye icon) and an 'Analyze' button (chart icon). The right sidebar shows details for the selected 'churn.csv' dataset, including its creation date, row and column counts, size, and available schemas.

Dataset Name	Rows	Columns
churn.csv	7043	21
Alteryxchurn.csv	7043	21
early_2012_2013_train-proc-reduced.csv	10000	30
loan-sampled20KRows.csv.zip	10000	74
soccerdemo.csv	17588	53
Yoga Studio Survey (1-Train).csv	1734	26
train.csv	41566	128
Kaggle_HRAnalytics.csv	1470	35
Airport Survey (1-Train).csv	5802	15
Auto Insurance Sales (All).csv	5802	15
UCI_Credit_Card.csv.zip	30000	25
Saas Customer Churn (1-Train).csv	1717	20
Combined Transaction Sheets 17-19 Items Out.csv	3477	61

**churn.csv Details:**

- Added: 11/12/2019
- Rows: 7043
- Columns: 21
- Size: 954.59 KB
- 1 Schemas: Default Schema

# Review Schema

Once all changes are made, save schema by clicking save as icon in the navigation bar.

You can also import and export schemas here.

You can exclude any columns that are overly unique, redundant or unnecessary information by clicking on the exclude box on the left. Excluded columns will highlight as red.

Review column data types. Click on the drop down arrows to change the data type.

Excluded	Pos	Column Name	Data Type	Uniques	Empties
<input checked="" type="checkbox"/>	1	customerID	Categorical	7043	0.00%
<input type="checkbox"/>	2	gender	Categorical	2	0.00%
<input type="checkbox"/>	3	SeniorCitizen	Categorical	2	0.00%
<input type="checkbox"/>	4	Partner	Categorical	2	0.00%
<input type="checkbox"/>	5	Dependents	Categorical	2	0.00%
<input type="checkbox"/>	6	tenure	Numeric	73	0.00%
<input checked="" type="checkbox"/>	7	PhoneService	Categorical	2	0.00%
<input type="checkbox"/>	8	MultipleLines	Categorical	3	0.00%
<input type="checkbox"/>	9	InternetService	Categorical	3	0.00%
<input type="checkbox"/>	10	OnlineSecurity	Categorical	3	0.00%
<input type="checkbox"/>	11	OnlineBackup	Categorical	3	0.00%
<input type="checkbox"/>	12	DeviceProtection	Categorical	3	0.00%
<input type="checkbox"/>	13	TechSupport	Categorical	3	0.00%
<input type="checkbox"/>	14	StreamingTV	Categorical	3	0.00%
<input type="checkbox"/>	15	StreamingMovies	Categorical	3	0.00%
<input type="checkbox"/>	16	Contract	Categorical	3	0.00%
<input type="checkbox"/>	17	PaperlessBilling	Categorical	2	0.00%
<input type="checkbox"/>	18	PaymentMethod	Categorical	4	0.00%
<input type="checkbox"/>	19	MonthlyCharges	Numeric	1585	0.00%



# Saving a Schema

You need to name your schema and, if needed, to add a more detailed description.

After your edits, click on the “Save” button.

SAVE SCHEMA

Name

Data Types Set

Description

changed column 11 from categorical to numeric

CANCEL

SAVE

# Run an Analysis

Select the dataset you wish to run your analysis on, it will highlight in blue.

Click on “Analyze” button in navigation bar.

The screenshot shows the Iris Pro interface with a grid of datasets. The 'churn.csv' dataset is highlighted in blue. The top navigation bar includes 'Review' and 'Analyze' buttons. The right sidebar shows details for the selected dataset, including 'Added', 'Rows', 'Columns', and 'Size'.

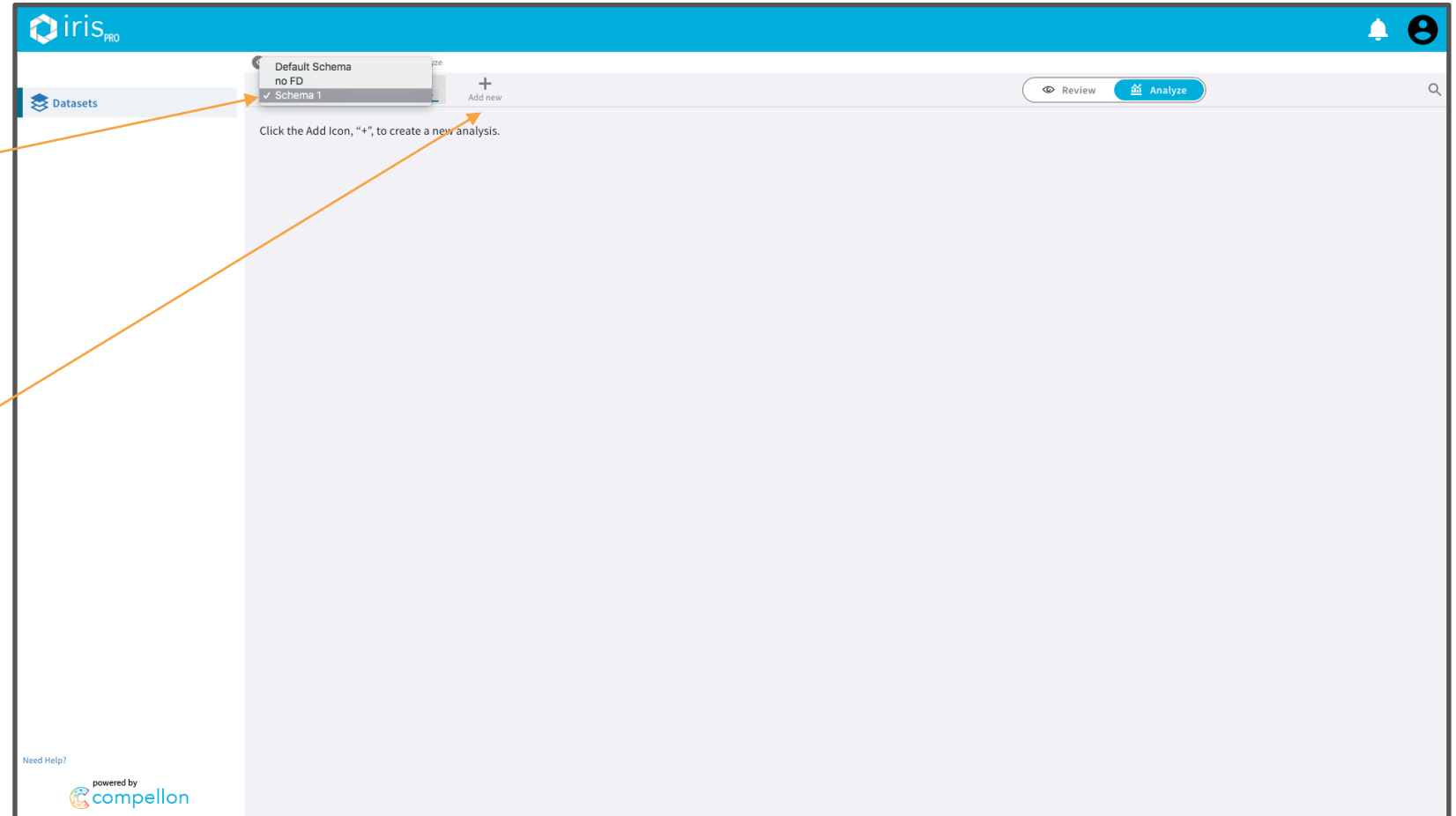
Dataset	Rows	Columns
churn.csv	7043	21
Alteryxchurn.csv	7043	21
early_2012_2013_train-proc-reduced.csv	10000	30
loan-sampled10KRows.csv	10000	74
soccerdemo.csv	17588	53
Yoga Studio Survey (1-Train).csv	1734	26
train.csv	41566	128
Kaggle_HRAnalytics.csv	1470	35
Airport Survey (1-Train).csv	5802	15
Auto Insurance Sales (All).csv	5802	15
UCI_Credit_Card.csv.zip	30000	25
Saas Customer Churn (1-Train).csv	1717	20
Combined Transaction Sheets 17-19 Items Out.csv	3477	61

Need Help? powered by compellon

# Run an Analysis

Select the schema you wish to apply from the Schema drop down list. If you didn't make any changes it will say "Default Schema"

Click on the plus sign to add new analysis.



# Run an Analysis

Name your analysis

Choose objective of analysis: Insights (historical, less stable) or Predict (forward looking, more stable)

Choose type of analysis: Targeted or Untargeted

Pick your target column and value. To add more than one click the plus sign.

Click on "Run"

The screenshot shows the 'ADD ANALYSIS' dialog box with the following elements and annotations:

- Analysis Name:** A text input field containing 'New Analysis 1'. An arrow points from the 'Name your analysis' instruction to this field.
- Objective:** A text block explaining that 'Insights' provides a historical view while 'Predict' provides more stable variables. Below this are two buttons: 'Insights' and 'Predict'. An arrow points from the 'Choose objective of analysis' instruction to the 'Predict' button.
- Type:** A text block asking 'Are you looking to target particular columns?'. Below it are 'Targeted' and 'Untargeted' buttons. An arrow points from the 'Choose type of analysis' instruction to the 'Targeted' button.
- Columns of Interest:** A dropdown menu showing 'Churn'. A plus sign button is below it. An arrow points from the 'Pick your target column and value' instruction to the 'Churn' dropdown.
- Values of Interest (Optional):** A text input field containing 'Yes'. An arrow points from the same instruction to this field.
- Buttons:** At the bottom right are 'CANCEL' and 'RUN' buttons. An arrow points from the 'Click on "Run"' instruction to the 'RUN' button.

# Review an Analysis

Analysis will show up on  
Analyze page.

Click on Name to see results

The screenshot displays the 'iris PRO' web application interface. At the top, a blue header bar contains the 'iris PRO' logo and a user profile icon. Below the header, a navigation bar shows 'Datasets' on the left and a search icon on the right. The main content area features a table with the following columns: 'Name', 'Targets', 'Run Date', and 'Run Time'. A single row is visible with the following data: 'Churn = Yes' (under Name), 'Churn = Yes' (under Targets), '11/12/2019 10:23:56 AM' (under Run Date), and '944 msec' (under Run Time). The 'Name' cell contains a blue link. An orange arrow points from the text 'Click on Name to see results' to this link. Above the table, there is a dropdown menu set to 'Schema 1' and an 'Add new' button. To the right of the table, there are 'Review' and 'Analyze' buttons. The bottom left corner of the interface includes a 'Need Help?' link and the 'powered by compellon' logo.

Name	Targets	Run Date	Run Time
<a href="#">Churn = Yes</a>	Churn = Yes	11/12/2019 10:23:56 AM	944 msec

# Review an Analysis

Here are the outputs of your analysis.

Click on the + to expand each segment.

Your identified target(s) will always be at the top.

The screenshot displays the Iris Pro interface for reviewing an analysis. The top navigation bar shows the file 'Auto Insurance Sales (All).csv'. The main content area is titled 'Sales Won - Predict' and lists several segments. The first segment, 'Target - Sale Won is Yes', is highlighted in red and has an orange arrow pointing to its expand icon (+). The other segments are listed below it, each with a plus icon (+) to expand it. The segments are: Target Candidate - Vehicle Class, Target Candidate - Marital Status, Target Candidate - Location Code, Target Candidate - Income, Target Candidate - EmploymentStatus, Information Relationship Graph, Non-Informative Columns, and Duplicate Columns. The sidebar on the left shows 'Datasets'. The bottom left corner has a 'Need Help?' link and the 'powered by compellon' logo.

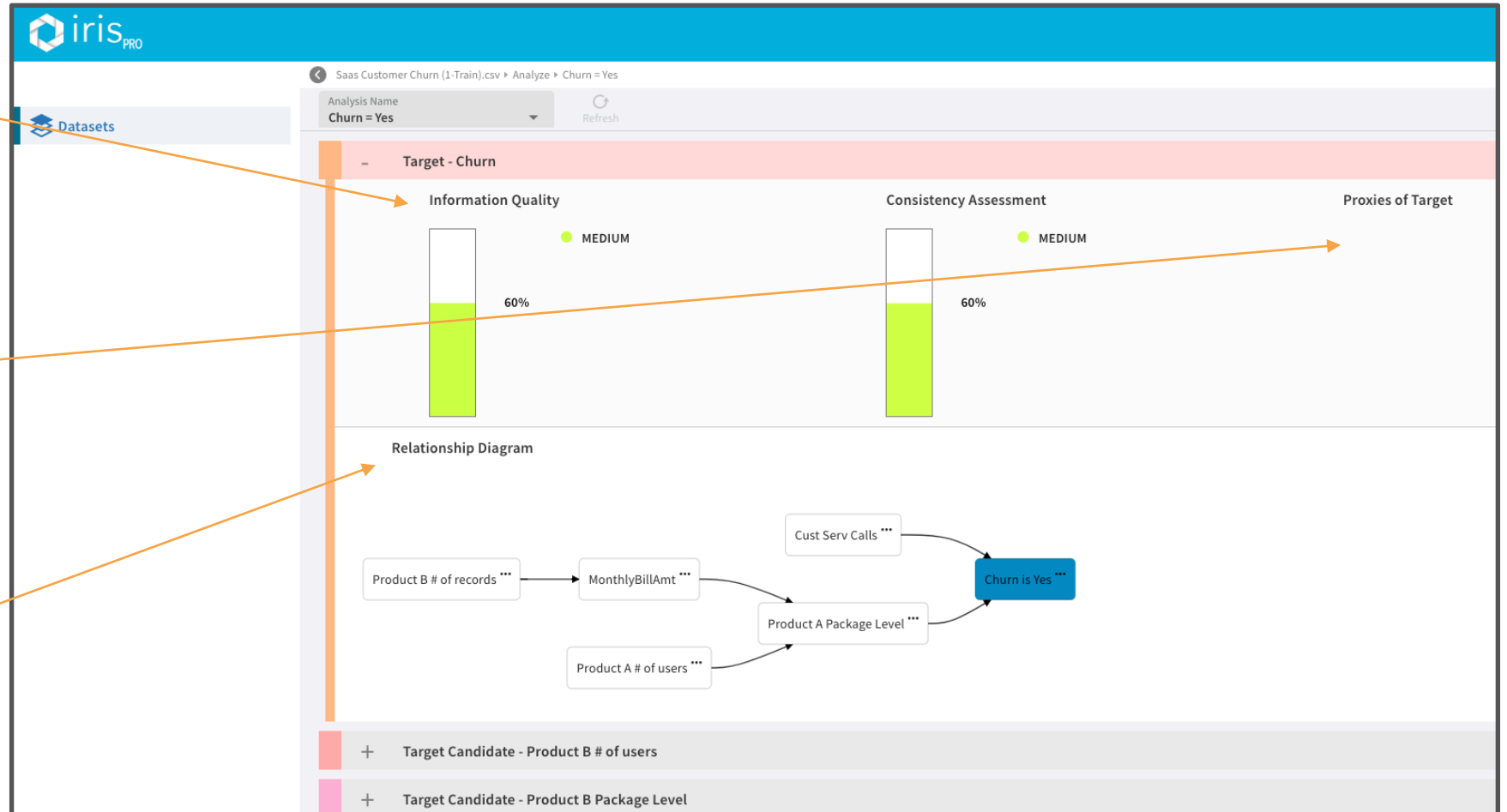
Segment
Target - Sale Won is Yes
Target Candidate - Vehicle Class
Target Candidate - Marital Status
Target Candidate - Location Code
Target Candidate - Income
Target Candidate - EmploymentStatus
Information Relationship Graph
Non-Informative Columns
Duplicate Columns

# Review an Analysis

For every user identified target or target candidate Iris Pro identifies measures of Information Quality and Consistency.

Any proxies of the target will be listed on the right

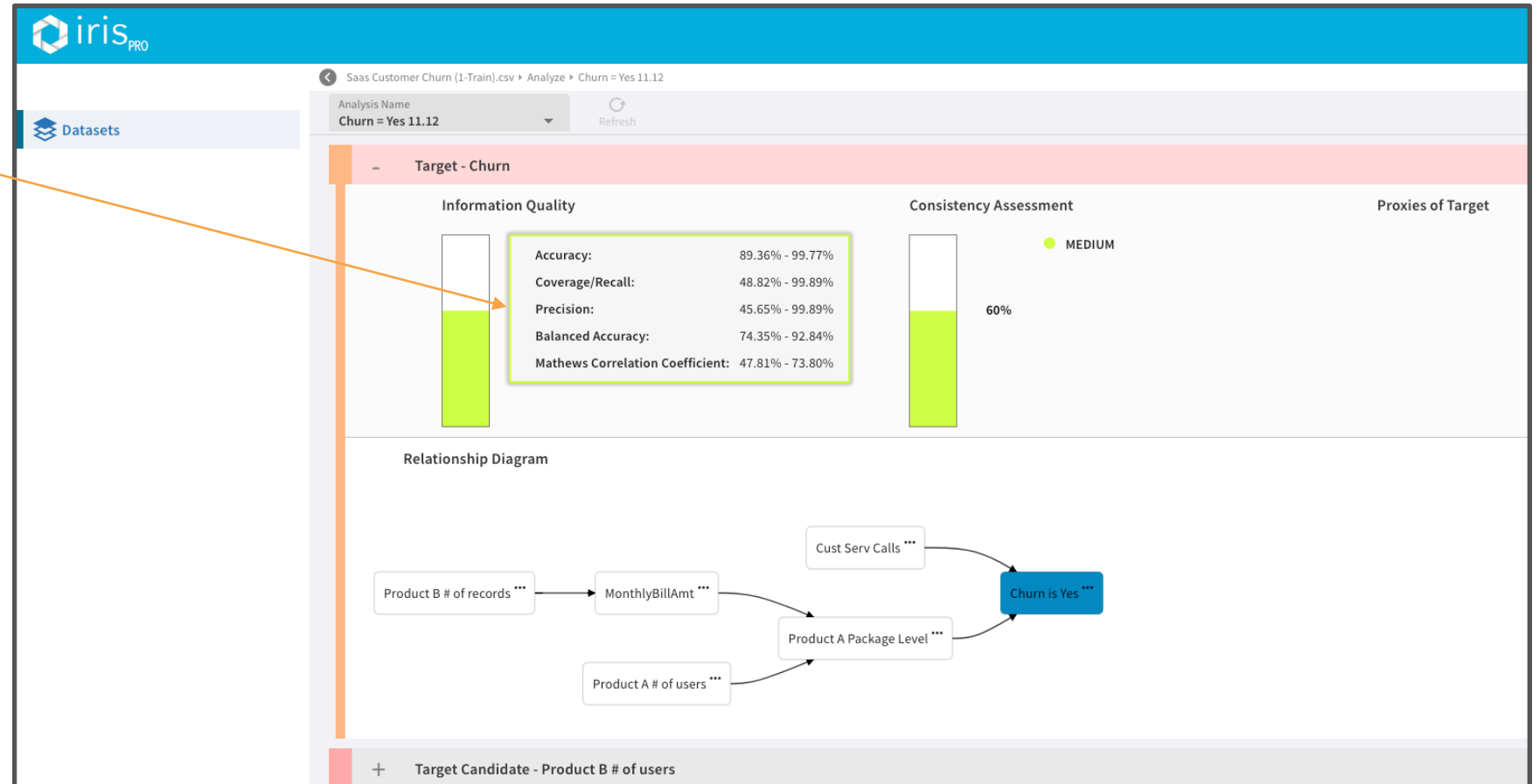
The best group of variables in your dataset that describe the target are represented in the Relationship Diagram.



# Review an Analysis

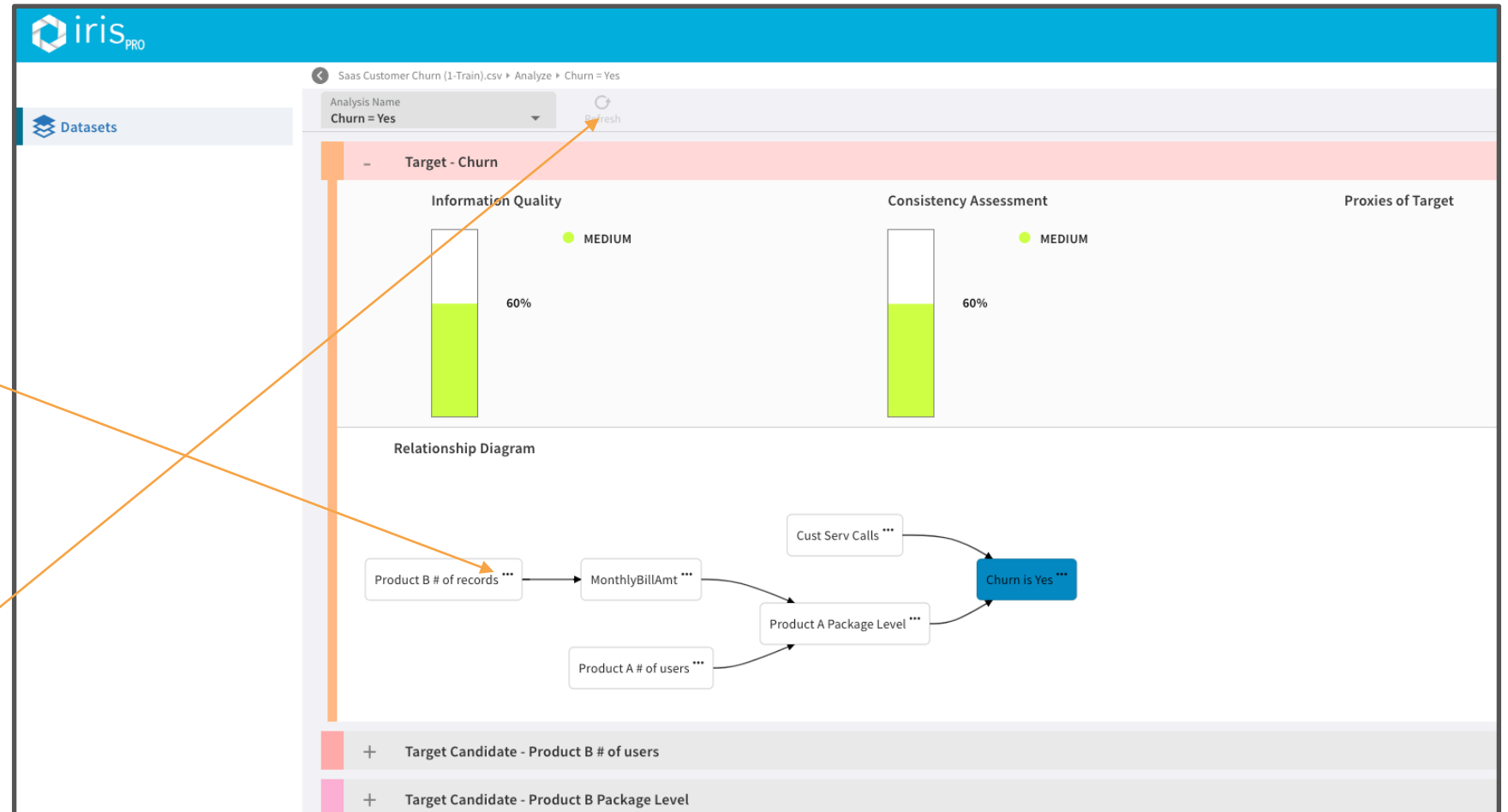
If you hover over the Information Quality bar the estimated model quality metrics will appear

These statistics give you the





# Review an Analysis



# Review an Analysis

The Relationship Diagram for the entire data set can be found under “Information Relationship Graph”

