These tools are new in 8.5 8.6 9.0 9.1 9.5

Favorites

Icon	Tool	Description	Example
	**Browse**	Add one or more points in your data stream to review and verify your data.	Allows users to get a look at their data anywhere in the process
	**Filter**	Query records based on an expression to split data into two streams, True (records that satisfy the expression) and False (those that do not).	If you are looking at a dataset of customers, you may want to only keep those customer records that have greater than 10 transactions or a certain level of sales.
	**Formula**	Create or update fields using one or more expressions to perform a broad variety of calculations and/or operations.	For instance if there is missing or NULL value, you can use the formula tool to replace that NULL value with a zero. Basically all of the formulas you can do in excel you can do inside Alteryx formula tool – so many more!
	**Input**	Bring data into your module by selecting a file or connecting to a database (optionally, using a query).	Connect to disparate datasets.
	**Join**	Combines two inputs based on a commonality between the two tables. Its function is like a SQL join but gives the option of creating 3 outputs resulting from the join.	Can be used to join Customer profile data as well as transactional data, and join the two data sources based on a unique customer ID
	**Output**	Output the contents of a data stream to a file or database.	Loading enriched data back into a database.
	**Sample**	Limit the data stream to a number, percentage, or random set of records.	Choosing the first 10 records for each region of previously sorted data so that you end up with the top 10 stores in each region.
	**Select**	Select, deselect, reorder and rename fields, change field type or size, and assign a description.	If your workflow only requires 5 fields out of 50 that are read in from the file/database, you can deselect all but the 5 required fields to speed up processing downstream.
	**Sort**	Sort records based on the values in one or more fields.	Allows you to sort data records into ascending/descending order- such as ranking your customers based on the amount of $$ they spend
	**Summarize**	Summarize data by grouping, summing, counting, spatial processing, string concatenation, and much more. The output contains only the results of the calculation(s).	You could determine how many customers you have in the state of NY and how much they have spent in total or an average per transaction.
	**Comment**	Add annotation or images to the module canvas to capture notes or explain processes for later reference.	This will allow uses to document what they did during a certain portion of the analysis, so other users have an understanding of what they were building in the workflow.
	**Text Input**	Manually add data which will be stored in the module.	A lookup table where you are looking for certain words or codes to be replaced with new classifications. You could create a Find field and a Replace field to populate with the values needed.
	**Union**	Combine two or more data streams with similar structures based on field names or positions. In the output, each column will contain the data from each input.	Transaction data stored in different files for different time periods, such as a sales data file for March and a separate one for April, can be combined into one data stream for further processing.

Input/Output

Icon	Tool	Description	Example
	**Browse**	Add one or more points in your data stream to review and verify your data.	Allows users to get a look at their data anywhere in the process

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Date Time Now**	Input the current date and time at module runtime, in a format of the user's choosing. (Useful for adding a date-time header to a report.)	This is a useful tool to easily add a date time header for a report
	**Directory**	Input a list of file names and attributes from a specified directory.	Lists all files in a directory- can be used in conjunction with the Dynamic Input tool to bring in the most recent data file that is available
	**Input**	Bring data into your module by selecting a file or connecting to a database (optionally, using a query).	Connect to disparate datasets.
	**Map Input**	Manually draw or select map objects (points, lines, and polygons) to be stored in the module.	Output results to google maps, and allow interaction with the results for example. Not really! This tool is only so you can pick a spatial object, either by drawing or selecting one, to use in your module (app).
	**Output**	Output the contents of a data stream to a file or database.	Output to data- anywhere we can read from, we can output to – no – we have some formats we can read only. Example- loading enriched data back into a database
	**Text Input**	Manually add data which will be stored in the module.	Manually store data or values inside the Alteryx module- example- a lookup table where you are values of segmentation groups and you want the description name
	**XDF Input**	This tool enables access to an XDF format file (the format used by Revolution R Enterprise's RevoScaleR system to scale predictive analytics to millions of records) for either: (1) using the XDF file as input to a predictive analytics tool or (2) reading the file into an Alteryx data stream for further data hygiene or blending activities	This can be used when building and running predictive analytics procedures on large amounts of data that open source R has difficulty computing - (specifically Linear Regression, Logistic Regression, Decision Trees, Random Forests, Scoring, Lift Chart)
	**XDF Output**	This tool reads an Alteryx data stream into an XDF format file, the file format used by Revolution R Enterprise's RevoScaleR system to scale predictive analytics to millions of records. By default, the new XDF files is stored as a temporary file, with the option of writing it to disk as a permanent file, which can be accessed in Alteryx using the XDF Input tool	This can be used when building and running predictive analytics procedures on large amounts of data that open source R has difficulty computing - (specifically Linear Regression, Logistic Regression, Decision Trees, Random Forests, Scoring, Lift Chart)

Preparation

Icon	Tool	Description	Example
	**Auto Field**	Automatically set the field type for each string field to the smallest possible size and type that will accommodate the data in each column.	Trying to identify the best fit field for text based inputs. Make the data streaming into Alteryx as small as possible to limit processing time and ensure proper formats for downstream processes.
	**Filter**	Query records based on an expression to split data into two streams, True (records that satisfy the expression) and False (those that do not).	Allows users to exclude values – all fields come through! in the stream- for instance if you are looking at a dataset of customers, you may want to eliminate certain characteristics of that customer such as race or sex
	**Date Filter**	The Date Filter macro is designed to allow a user to easily filter data based on a date criteria using a calendar based interface.	Return transaction records by specifying a start and end date.
	**Formula**	Create or update fields using one or more expressions to perform a broad variety of calculations and/or operations.	For instance if there is missing or NULL value, you can use the formula tool to replace that NULL value with a zero

These tools are new in 8.5 8.6 9.0 9.1 9.5

Icon	Tool	Description	Example
	**Generate Rows**	Create new rows of data. Useful for creating a sequence of numbers, transactions, or dates.	Creating data specifically time series data, Create 365 unique records for each day of year.
	**Impute Values**	Update specific values in a numeric data field with another selected value. Useful for replacing NULL() values.	For example, if you have a data set that is missing information, such as salary, and displays (NULL) rather than just making it zero, you can use the mean or median to fill in the NULL, to improve accuracy of the results.
	**Multi-Field Binning**	Group multiple numeric fields into tiles or bins, especially for use in predictive analysis.	For instance if you have transactional data, you can group them into different buyer personas- ie.. Males between 30-35, that spend between $1k> per month, etc…
	**Multi-Field Formula**	Create or update multiple fields using a single expression to perform a broad variety of calculations and/or operations.	For instance if there is missing or NULL value on multiple fields, you can use the formula tool to replace that NULL value with a zero
	**Multi-Row Formula**	Create or update a single field using an expression that can reference fields in subsequent and/or prior rows to perform a broad variety of calculations and/or operations. Useful for parsing complex data and creating running totals.	Creating unique identifiers at a group level- cross row comparisons- Sales volume for yr1, yr2, yr3 in different rows and want to notice the difference between the sales in each of these rows
	**Random % Sample**	Generate a random number or percentage of records passing through the data stream.	If you want to base your analysis based on 35% of the data for instance, it will randomly return records
	**Record ID**	Assign a unique identifier to each record.	This can be used to assign a customer id to a legacy transaction, allowing for more accurate direct marketing/promotional offerings in the future
	**Sample**	Limit the data stream to a number, percentage, or random set of records.	Allows you to select a subset of data/records for your analysis- Can be used to focus on a select group of related records or transactions for analysis- such as selecting all items in an online shopping cart
	**Select**	Select, deselect, reorder and rename fields, change field type or size, and assign a description.	Allows you to determine a specific subset of records should or should not be carried down throughout the analysis- for instance if we are looking at customer transactional data and we want to eliminate all transactions that are less than $5K)
	**Select Records**	Select specific records and/or ranges of records including discontinuous ranges. Useful for troubleshooting and sampling.	If a user wants to find records that are less than $100 or in a range of $100-$150 it will return records in this range
	**Sort**	Sort records based on the values in one or more fields.	Allows you to sort data records into ascending/descending order- such as locating your top 1000 customers based on the amount of $$ they spend
	**Tile**	Group data into sets (tiles) based on value ranges in a field.	Creating logical groups of your data. User defined breaks or statistical breaks. Very good for bucketing high valued customers vs. low valued customers
	**Unique**	Separate data into two streams, duplicate and unique records, based on the fields of the user's choosing.	Only want to mail to one individual, and based on a unique identifier(customer id)
Join			
Icon	**Tool**	**Description**	**Example**
	**Append Field**	Append the fields from a source input to every record of a target input. Each record of the target input will be duplicated for every record in the source input.	Adding small value to a million records.. A small to big merge- Adding time stamps as well as the name of the users who last accessed it onto your database records.

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Find Replace**	Search for data in one field from one data stream and replace it with a specified field from a different stream. Similar to an Excel VLOOKUP.	Think of this like Excel- find and replace. Looking for something and then replacing.
	**Join**	Combine two data streams based on common fields (or record position). In the joined output, each row will contain the data from both inputs.	For instance this can be used to join Customer profile data as well as transactional data, and join the two data sources based on a unique customer ID
	**Join Multiple**	Combine two or more inputs based on common fields (or record position). In the joined output, each row will contain the data from both inputs.	For instance this can be used to join Customer profile data as well as transactional data, and join two or more data sources based on a unique customer ID
	**Make Group**	The Make Group tool takes data relationships and assembles the data into groups based on those relationships.	Used primarily with Fuzzy Matching- ID 1 can match 10 different values from source 2 and that becomes a group.
	**Fuzzy Match**	Identify non-identical duplicates in a data stream.	Helps determine similarities in your data. For instance if you have 2 different data sets with different ID, looking a names and address as a way to standard and matching them up based on these types of characteristics- and displays all of the id's that match
	**Business Listing Matching**	Match your customer or prospect file to the Dun & Bradstreet business file.	Matching a business listing file to Dun and Bradstreet.
	**Household Matching**	Match your customer file to the Experian Consumer View Household file.	Matching a customer file to Experian- example to append segmentation values at a household level
	**Union**	Combine two or more data streams with similar structures based on field names or positions. In the output, each column will contain the data from each input.	Can be used to combine datasets with similar structures, but with different data. You might have transaction data stored in different files for different time periods, such as a sales data file for March and a separate one for April. Assuming that they have the same structure (the same fields), Union will join them together into one large file, which you can then analyze

Parse

Icon	Tool	Description	Example
	**Date Time**	Transform date/time data to and from a variety of formats, including both expression-friendly and human readable formats.	Easy conversion between strings and actual date time formats- Example- taking Military time into standard times. Or turning Jan 1, 2012 into 1.1.12, etc..
	**RegEx**	Parse, match, or replace data using regular expression syntax.	An example would be if someone is trying to parse unstructured text based files- Weblogs or data feeds from twitter, helps arrange the data for analytical purposes into rows and columns.
	**Text to Columns**	Split the text from one field into separate rows or columns.	Allows you bring customer data from an excel file for instance that contains first name and last name in one column and split them up into 2 columns so first name is in one column and last name is in the other column- this will make it easy to sort and analyze
	**XML Parse**	Read in XML snippets and parse them into individual fields.	Cleaning an xml file, parse xml text

Transform

Icon	Tool	Description	Example
	**Arrange**	Manually transpose and rearrange fields for presentation purposes.	Used for staging data for reports

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Count Records**	Count the records passing through the data stream. A count of zero is returned if no records pass through.	Returns a count of how many records are going through the tool
	**Cross Tab**	Pivot the orientation of the data stream so that vertical fields are on the horizontal axis, summarized where specified.	Think of it as a way to change your excel spreadsheet that has a column of customer ID's and then next to a column of revenue. This will turn these two columns into two rows
	**Running Total**	Calculate a cumulative sum per record in a data stream.	Can take 3 columns of sales totals and summarize 3 yr totals of sales by that row.. (ie yr 1 sales 10K, yr 2 15K, yr 3 25K)
	**Summarize**	Summarize data by grouping, summing, counting, spatial processing, string concatenation, and much more. The output contains only the results of the calculation(s).	For instance if you wanted to look at a certain group of customers of a certain age or income level, or get an idea of how many customers you have in the state of NY
	**Transpose**	Pivot the orientation of the data stream so that horizontal fields are on the vertical axis.	Think of it as a way to change your excel spreadsheet that has a row of customer ID's and then below that a row of revenue. This will turn these two rows into two columns
	**Weighted Average**	Calculate the weighted average of a set of values where some records are configured to contribute more than others.	So if you are looking at calculating average spend, this will determine and "weight" if there are certain customers spending levels that are contributing to the average.

Report / Presentation

Icon	Tool	Description	Example
	**Charting**	Create a chart (Area, Column, Bar, Line, Pie, etc.) for output via the Render tool.	Create bar, line, pie charts
	**Email**	Send emails for each record with attachments or e-mail generated reports if desired.	Allows you to create dynamically updated email content
	**Image**	Add an image for output via the Render tool.	Add graphics/image that will be included in report
	**Layout**	Arrange two or more reporting snippets horizontally or vertically for output via the Render tool.	How to arrange the pieces of your report
	**Report Map**	Create a map for output via the Render tool.	Create a map for a report
	**Map Legend Builder**	Recombine the component parts of a map legend (created using the Map Legend Splitter) into a single legend table, after customization by other tools.	Takes a customized legend and reassembles it.
	**Map Legend Splitter**	Split the legend from the Report Map tool into its component parts for customization by other tools. (Generally recombined by the Map Legend Builder.)	Help customize legends by adding symbols such as $ or % for instance or removing redundant text
	**Overlay**	Arrange reporting snippets on top of one another for output via the Render tool.	Allows you to specify how to put a map together- example, putting a legend inside a map – I'd suggest a different example such as…overlaying a table and chart onto a map
	**Render**	Output report snippets into presentation-quality reports in a variety of formats, including PDF, HTML, XLSX and DOCX.	Saves reports out of Alteryx

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Report Footer**	Add a footer to a report for output via the Render tool.	Apply a footer to the report
	**Report Header**	Add a header to a report for output via the Render tool.	Apply a header to the report
	**Table**	Create a data table for output via the Render tool.	Creates table for selected data fields
	**Report Text**	Add and customize text for output via the Render tool.	Allows you to customize a title or other text related aspects to your report

Documentation

Icon	Tool	Description	Example
	**Comment**	Add annotation or images to the module canvas to capture notes or explain processes for later reference.	This will allow users to document what they did during a certain portion of the analysis, so other users have an understanding
	**Explorer Box**	Add a web page or Windows Explorer window to your canvas.	Helps you organize URL embedded into the canvas – I suggest a different example: Display a web page for reference in the module or use it to show a shared directory of macros
	**Tool Container**	Organize tools into a single box which can be collapsed or disabled.	Helps you organize your module.

Spatial

Icon	Tool	Description	Example
	**Buffer**	Expand or contract the extents of a spatial object (typically a polygon).	Identify all of the business on a road, by placing a buffer on that road to determine who they are/where they are
	**Create Points**	Create spatial points in the data stream using numeric coordinate fields.	Finding a spatial ref to a longitude, latitude
	**Distance**	Calculate the distance or drivetime between a point and another point, line, or polygon.	Creating the drive distance or drive time to a customer location
	**Find Nearest**	Identify the closest points or polygons in one file to the points in a second file.	As a customer, find me the nearest location to visit, optimizing my driving route
	**Generalize**	Simplify a polygon or polyline object by decreasing the number of nodes.	Object processing- generating an output map of a coastal boundary and you don't need it to be too detailed – significantly decrease the physical size of a spatial record can increase processing time
	**Heat Map**	Generate polygons representing different levels of "heat" (e.g. demand) in a given area, based on individual records (e.g. customers)	Could be used to view where there is a heavier amount of households in a certain location
	**Make Grid**	Create a grid within spatial objects in the data stream.	Use it to bucket on the ground where their customers are coming from (area, etc..)
	**Non Overlap Drivetime**	Create drivetime trade areas that do not overlap for a point file.	Create drivetime trade areas that do not overlap, for a point file

These tools are new in 8.5 8.6 9.0 9.1 9.5

Icon	Tool	Description	Example
	**Poly-Build**	Create a polygon or polyline from sets of points.	Build a trade area- build an object of where all of my customers are coming from. Building a polygon to fit a series of points
	**Poly-Split**	Split a polygon or polyline into its component polygons, lines, or points.	Break a polygon into a sequential set of points. The results can be rendered in Tableau? I'll take your word for it, but this one doesn't make sense to me (Tableau part)
	**Smooth**	Round off sharp angles of a polygon or polyline by adding nodes along its lines.	Crisp objects rendered on a map. (coastal view more detailed)
	**Spatial Info**	Extract information about a spatial object, such as area, centroid, bounding rectangle, etc.	Getting the Lat/Lon of a point. Maybe the area square miles of a cover area for telco/wireless
	**Spatial Match**	Combine two data streams based on the relationship between two sets of spatial objects to determine if the objects intersect, contain or touch one another.	Finding all customers that fall within a defined trade area, based on their geographic proximity
	**Spatial Process**	Create a new spatial object from the combination or intersection of two spatial objects.	Want to remove overlap from intersecting trade areas.
	**Trade Area**	Define radii (including non-overlapping) or drive-time polygons around specified points.	Defining boundaries for where your customers or prospects are coming from
Data Investigation			
Icon	**Tool**	**Description**	**Example**
	**Association Analysis**	Determine which fields in a database have a bivariate association with one another.	For example, if the user is trying to determine who should be contacted as part of a direct marketing campaign, to estimate the probability a prospect will respond favorably if contacted in the marketing campaign.
	**Contingency Table**	Create a contingency table based on selected fields, to list all combinations of the field values with frequency and percent columns.	For example you can build a table of males and females and how many times they purchase certain products during a weeks time.
	**Create Samples**	Split the data stream into two or three random samples with a specified percentage of records in the estimation and validation samples. If the total is less than 100%, the remaining records fall in the holdout sample.	For example, in the case of a direct marketing campaign, we want to know the probability that a prospect that is contacted as part of the campaign will respond favorably to it, *before* we include that prospect on the contact list of the campaign in order to make the decision whether to include the prospect on the list. As a result, what we really care about in selecting a predictive model to implement a business process is the ability of that model to accurately predict new data (the model that does the best job of predicting data in the estimation sample does not do the best job of predicting new data since it "overfits" the estimation sample). To do this, we need to know the actual outcome in order to assess model accuracy. As a result, we will take data where we know the outcomes (perhaps as a result of a test implementation of that campaign), and use part of the data (the estimation sample) to create a set of candidate predictive models, and another, separate, part of the data (the validation sample) to compare the ability of the different candidate models to predict the outcomes for this second set of data in order to select the model to put into production. At times the user may want to use a third portion of the available data (the holdout sample) for the purposes of developing unbiased estimates of the economic implications of putting a model into a production business process.

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Distributed Analysis**	Allows you to fit one or more distributions to the input data and compare them based on a number of Goodness-of-Fit* statistics. Based on the statistical significance (p-values) of the results of these tests, the user can determine which distribution best represents the data.	Helpful when trying to understand the overall nature of your data as well as make decisions about how to analyze it. For instance, data that fits a Normal distribution would likely be well-suited to a Linear Regression, while data that is Gamma Distributed be better-suited to analysis via the Gamma Regression tool.
	**Field Summary Report**	Produce a concise summary report of descriptive statistics for the selected data fields.	This tool provides a concise, high level overview of all the fields in a database. This information can be invaluable to users in determining what fields they need to pay special attention to in their analysis. For instance, if State is a field in a customer database for an online retailer, and there is only a single customer from the state of Alaska, the analyst will quickly be able to determine that any analysis (ranging from simple means by state to more advanced predictive models) that involve the State field will result in very unreliable information for Alaska. Given this information, the user may decide to not use the State field, or to combine the Alaska customer with customers from another state (perhaps Hawaii) for analysis purposes.
	**Frequency Table**	Produce a frequency analysis for selected fields - output includes a summary of the selected field(s) with frequency counts and percentages for each value in a field.	For example, What is the distribution of a company's customers by income level? From the output, you might learn that 35% of your customers are in high income, 30% are in middle high, 25% are in middle low, and 10% are in low.
	**Histogram**	Provides a histogram plot for a numeric field. Optionally, it provides a smoothed empirical density plot. Frequencies are displayed when a density plot is not selected, and probabilities when this option is selected. The number of breaks can be set by the user, or determined automatically using the method of Sturges.	Provides a visual summary of the distribution of values based on the frequency of intervals. For example, the US Census using their data on the time occupied by travel to work, Table 2 below shows the absolute number of people who responded with travel times "at least 15 but less than 20 minutes" is higher than the numbers for the categories above and below it. This is likely due to people rounding their reported journey time.
	**Heat Plot**	This tools plots the empirical bivariate density of two numeric fields using colors to indicate variations in the density of the data for different levels of the two fields	Could be used for reports to visualize contingency tables for media usage. For example, a survey of how important Internet reviews were in making a purchase decision (on a 1 to 10 scale), and when the customer searched for this information (done in time categories from six or more from the time of purchase to within one hour of the point of purchase). The heat plot allows user to see that those who looked at internet reviews five to six weeks from the point of purchase were most influenced by those reviews.
	**Oversample Field**	Sample incoming data so that there is equal representation of data values to enable effective use in a predictive model.	In many applications, the behavior of interest (e.g., responding favorably to a promotion offer) is a fairly rare event (untargeted or poorly targeted direct marketing campaigns often have favorable response rates that are below 2%). Building predictive models directly with this sort of rare event data is a problem since models that predict that no one will respond favorably to a promotion offer will be correct in the vast majority of cases. To prevent this from happening, it is common practice to oversample the favorable responses so that there is a higher penalty for placing all customers into a non-responder category. This is done by taking all the favorable responders and a sample of the non-responders to get the total percentage of favorable responders up to a user specified percentage (often 50% of the sample used to create a model).
	**Pearson Correlation**	Replaces the Pearson Correlation Coefficient in previous versions… The Pearson coefficient is obtained by dividing the covariance of the two variables by the product of their standard deviations	For example age and income are related. So as age increases so will income.

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Plot of Means**	Take a numeric or binary categorical (converted into a set of zero and one values) field as a response field along with a categorical field and plot the mean of the response field for each of the categories (levels) of the categorical field.	The best use of this tool is for gaining a basic understanding of the nature of the relationship between a categorical variable and numeric variable. For instance, it allows us to visually examine whether customers in different regions of the country spend more or less on women's apparel in a year.
	**Scatterplot**	Produce enhanced scatterplots, with options to include boxplots in the margins, a linear regression line, a smooth curve via non-parametric regression, a smoothed conditional spread, outlier identification, and a regression line. The smooth curve can expose the relationship between two variables relative to a traditional scatter plot, particularly in cases with many observations or a high level of dispersion in the data.	Shows the relationship between two numeric variables or a numeric variable and a binary categorical variable (e.g., Yes/No). In addition to the points themselves, the tool also produces lines that show the trends in the relationships. For instance, it may show us that household spending on restaurant meals increases with household income, but the rate of increase slows (i.e., shows "diminishing returns") as the level of household income increases.
	**Spearman Correlation Coefficient**	Assesses how well an arbitrary monotonic function could describe the relationship between two variables without making any other assumptions about the particular nature of the relationship between the variables.	For example, is there a correlation between income level and their level of education
	**Violin Plot**	Shows the distribution of a single numeric variable, and conveys the density of the distribution based on a kernel smoother that indicates the density of values (via width) of the numeric field. In addition to concisely showing the nature of the distribution of a numeric variable, violin plots are an excellent way of visualizing the relationship between a numeric and categorical variable by creating a separate violin plot for each value of the categorical variable.	For example, it can create a plot of the distribution of the number of minutes of cell phone talk time used by different customer age group categories in a particular month. In this way, the tool allows a data artisan to gain a more complete understanding of a particular field, or of the relationship between two different fields (one categorical and one numeric).

Predictive

Icon	Tool	Description	Example
	**AB Test Analysis**	Compare the percentage change in a performance measure to the same measure one year prior.	For example comparing Tuesday lunch traffic at a restaurant last year (when the test was not run) to Tuesday lunch traffic for the same week this year (when the test was run).
	**AB Controls**	The Control Select tool matches one to ten control units (e.g., stores, customers, etc.) to each member of a set of previously selected test units, on the criteria such as seasonal patterns and growth trends for a key performance indicator, along with other user provided criteria.	AB Controls takes the two things like (seasonality, growth, etc.) and for the treatment stores, compares them versus the control candidates (the other stores in the chain) that are nearest to those stores (seasonality, growth) within a given drive time. – so compare low-growth store to other low-growth stores within X distance. The goal is to find the best set of control units (those units that did not receive the test treatment, but are very similar to a unit that did on important criteria) for the purposes of doing the best comparison possible.
	**AB Treatments**	Determine which group is the best fit for AB testing.	For example choosing which DMA that you would want to compare using up to 5 criteria that you want to use at the treatment observation level (minimal criteria, as well as the spread between criteria between DMAs and within DMAs based on the criteria that you want). Absolute comparison would be against the average for the entire chain / customer set.
	**AB Trends**	Create measures of trend and seasonal patterns that can be used in helping to match treatment to control units (e.g., stores or customers) for A/B testing. The trend measure is based on period to period percentage changes in the rolling average	For example, it gives users the ability to cluster treatment observations units based on underlying trends over the course of a year (general growth rate month-to-month = high-growth, low-growth, medium growth), or specific seasonality patterns (climate zones,) based on your measures (like traffic, sales

These tools are new in 8.5 8.6 9.0 9.1 9.5

		(taken over a one year period) in a performance measure of interest. The same measure is used to assess seasonal effects. In particular, the percentage of the total level of the measure in each reporting period is used to assess seasonal patterns.	volumes) and frequency (daily, weekly, monthly) and over what period of time. Can do day specific data (Mondays versus Tuesdays versus etcetera).
	**Boosted Model**	Provides generalized boosted regression models based on the gradient boosting methods of Friedman.* It works by serially adding simple decision tree models to a model ensemble so as to minimize an appropriate loss function.	Provides a visual output that enables an understanding of both the relative importance of different predictor fields on the target, and the nature of the relationship between the target field and each of the important predictor fields. Such as most important variables related to churn or which variables to focus on in a targeted campaign.
	**Count Regression**	Estimate regression models for count data (e.g., the number of store visits a customer makes in a year), using Poisson regression, quasi-Poisson regression, or negative binomial regression. The R functions used to accomplish this are glm() (from the R stats package) and glm.nb() (from the MASS package).	Regression models for count data (e.g., integer values like the number of numbers to a cell phone account, the number of visits a customer makes to our store in a given year) that are integer in nature. Like linear/logistic regression; typically using with small numbers (visits to a doctor's office – always a positive number, and typically an integer value) – helps address the risk of biased results in linear regression where you have a relatively small number of possible positive integer values
	**Decision Tree**	Predict a target variable using one or more predictor variables that are expected to have an influence on the target variable by constructing a set of if-then split rules that optimize a criteria. If the target variable identifies membership in one of a set of categories, a classification tree is constructed (based on Gini coefficient) to maximize the 'purity' at each split.If the target variable is a continuous variable, a regression tree is constructed using the split criteria of 'minimize the sum of the squared errors' at each split.	A decision tree creates a set of if-then rules for classifying records (e.g., customers, prospect, etc.) into groups based on the target field (the field we want to predict). For instance, in the case of evaluating a credit union's applicants for personal loans, the credit union can use the method with data on past loans it has issued and find that customers who had: (a) an average monthly checking balances of over $1,500 (b) no outstanding personal loans (c) were between the ages of 50 and 59 had a default rate of less than 0.3%, and therefore should have their loan applications approved.
	**Forest Model**	Predict a target variable using one or more predictor variables that are expected to have an influence on the target variable, by constructing and combining a set of decision tree models (an "ensemble" of decision tree models).	The forest model is actually a group of decision tree models (the group is typically called an "ensemble"), where the trees within the group differ with respect to the set of records used to create the decision trees and the set of variables that were considered in creating the if-then rules in the tree. Prediction in this model are based on the predictions of all the models, with the final prediction on being the outcome the greatest number of models in the group predicted (this is known as a "voting rule").
	**Gamma Regression**	based on the R and Revo generalized linear model, called the Gamma Regression, which is based on an underlying Gamma distribution) that handles strictly positive target variables that have a long right tail (so most values are relatively small, and there is a long right-hand tail to the distribution)	It is a more appropriate model to use than linear regression when the target is strictly positive and has a long right-hand tail (so a small number of customers with high values of the target, and lots with relatively low values). It is widely used in the insurance industry to model claim amounts (where most claims are fairly small, but some can be very large).
	**Lift Chart**	Compare the improvement (or lift) that various models provide to each other as well as a 'random guess' to help determine which model is 'best.'Produce a cumulative captured response chart (also called a gains chart) or an incremental response rate chart.	A cumulative captured response lift chart can tell us that the best 10% of customers (based on a predictive model) in our customer list for a direct marketing campaign resulted in 40% of the total favorable response we would get if all of the members of our customer list were contacted as part of the campaign. In addition, an incremental response rate chart can tell us that the second best 10% of our customer list based on a predictive model (so customers that were ranked just under the best 10% to the last customer ranked in the top 20%) have a favorable response rate of 8% compared to the overall response rate of 2%.

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Market Basket Rules**	Step 1 of a Market Basket Analysis: Take transaction oriented data and create either a set of association rules or frequent item sets. A summary report of both the transaction data and the rules/item sets is produced, along with a model object that can be further investigated in an MB Inspect tool.	For example the Market basket rule can state that is someone purchases Beer they are most likely to also purchase pizza as well or if they purchase Fish, they are most likely to purchase white wine at the same time.
	**Market Basket Inspect**	Step 2 of a Market Basket Analysis: Take the output of the MB Rules tool, and provide a listing and analysis of those rules that can be filtered on several criteria in order to reduce the number or returned rules or item sets to a manageable number.	A tool for inspecting and analyzing association rules and frequent items sets. So do the Beer and Pizza MB Rules really fit? • Filter association rules and review them graphically (two different visuals) • Can look at certain levels of support/confidence/lift • Used to fine tune rules (which ones to use/keep/focus on) • What are the rules that begin to make sense --> outputs the rules (yxdb stream)
	**Nested Test**	Examine whether two models, one of which contains a subset of the variables contained in the other, are statistically equivalent in terms of their predictive capability.	This tool allows the user to determine whether one or more fields help to predict the target field for a regression based model in the presence of other predictor fields. By using this tool, a user can find a statistically meaningful set of predictors for the target field in a very selective way, as opposed to the automated method that the Regression – Stepwise tool represents.
	**Linear Regression**	Relate a variable of interest (target variable) to one or more variables (predictor variables) that are expected to have an influence on the target variable. (Also known as a linear model or a least-squares regression.)	For instance is the number of times someone shops at a store related to their income level
	**Logistic Regression**	Relate a binary (yes/no) variable of interest (target variable) to one or more variables (predictor variables) that are expected to have an influence on the target variable.	For instance what is the probability that someone who graduated five years ago in engineering from a university will make a donation to that university if they are included in a telephone based fund raising campaign? How does this probability compare to the probability that someone who graduate from the university 20 years ago with a degree in education will donate to the same campaign (i.e., which of these two people represents a better donation prospect)?
	**Naives Bayes**	Creates a binomial or multinomial probabilistic classification model of the relationship between a set of predictor variables and a categorical target variable. The Naive Bayes classifier assumes that all predictor variables are independent of one another and predicts, based on a sample input, a probability distribution over a set of classes, thus calculating the probability of belonging to each class of the target variable.	Predicting whether someone leasing a new vehicle will purchase that car at the termination of the lease based on both the characteristics of the vehicle or (e.g., pickup/sedan/SUV) and the customer (e.g., gender, age, etc.).
	**Neural Networks**	This tool allows a user to create a feedforward perceptron neural network model with a single hidden layer. The neurons in the hidden layer use a logistic (also known as a sigmoid) activation function, and the output activation function depends on the nature of the target field. Specifically, for binary classification problems (e.g., the probability a customer buys or does not buy), the output activation function used is logistic, for multinomial classification problems (e.g., the probability a customer chooses option A, B, or C) the output activation function used is softmax, for regression problems (where the target is a continuous, numeric field) a linear activation function is used for the output.	Can be used to help in financial risk assessment by scoring an applicant to determine the risk in extending credit or detect fraudulent transactions in an insurance claims database.

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Support Vector Machine**	Support Vector Machines (SVM), or Support Vector Networks (SVN), are popular supervised learning algorithms used for classification problems, and are meant to accommodate instances where the data (i.e., observations) are considered linearly non-separable. In other words, the target values cannot be separated into their underlying classes using a simple, single linear boundary.*	It can be used for both classification and regression models. It has become a "go to" tool for a lot of machine learning applications due to its ability to scale very well. So it can be used in areas like churn or fraud detection for example.
	**Spline Model**	This tool implements Friedman's multivariate adaptive regression spline (MARS) model. This is in the more modern class of models (like the Forest and Boosted Models) that handles both variable selection and non-linear relationships directly with the algorithm. In some ways it is similar to a decision tree, but instead of making discrete jumps at "splits", the splits (called "knots" in this method) place in a "hinge", where the slope of the effect of a predictor on a target changes, resulting in the effect of numeric predictors being modeled as piecewise linear components	Similar to Forest Model or Boosted Model it provides a visual output that enables an understanding of both the relative importance of different predictor fields on the target, and the nature of the relationship between the target field and each of the important predictor fields. Such as most important variables related to churn or which variables to focus on in a targeted campaign.
	**Stepwise**	Determine the "best" predictor variables to include in a model out of a larger set of potential predictor variables for linear, logistic, and other traditional regression models. The Alteryx R-based stepwise regression tool makes use of both backward variable selection and mixed backward and forward variable selection.	This tool takes a linear or logistic regression model (likely one that includes a large number of predictor fields) and finds the subset of the predictors that provides the best adjusted (for the number of model parameters) model fit. In general, a model with fewer predictors is better at predicting new data than one with more predictors since it is less prone to over fitting the estimation sample.
	**Score**	Calculate a predicted value for the target variable in the model. This is done by appending a 'Score' field to each record in the output of the data stream, based on the inputs: an R model object (produced by the Logistic Regression, Decision Tree, Forest Model, or Linear Regression) and a data stream consistent with the model object (in terms of field names and the field types).	This tool takes a model and provides predicted model values for the target variable in new data. This is what actually enables a predictive model to be incorporated into a business process.
	**Test of Means**	Compare the difference in mean values (using a Welch two sample t-test) for a numeric response field between a control group and one or more treatment groups.	Used (but not exclusively) in AB testing, compares two groups across any given measures based on their mean value. Does the average person in AZ buy as much as the average in Maine? Is that difference significant or could it be random chance.

Time Series

Icon	Tool	Description	Example
	**TS ARIMA**	Estimate a univariate time series forecasting model using an autoregressive integrated moving average (or ARIMA) method.	The two most commonly used methods of univariate (single variable) time series forecasting tools are ARIMA and exponential smoothing. This tool implements the ARIMA model, and can be run in a fully automated way. In other words, the user provides a single time series; say monthly sales for a particular product, and the tool figures out the best ARIMA model given the past data for forecasting future sales levels of that product. Alternatively, experienced users have the ability to alter any of the parameter setting used to create the model
	**TS Compare**	Compare one or more univariate time series models created with either the ETS or ARIMA tools.	This tool allows two or more time series forecasting models (based on models created using the TS ARIMA and TS ETS tools) to be compared in terms of their predictive accuracy. This enables the analyst to determine the best model to use for making future forecasts of values of interest (say monthly sales of a product).

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**TS ETS**	Estimate a univariate time series forecasting model using an exponential smoothing method.	This tool implements the creation of univariate time series models using exponential smoothing. As with the TS ARIMA tool, the user can run the tool in a fully automated mode or alter any of the method parameters used in creating models.It can help you understand the effect that factors such as economic and market conditions, customer demographics, pricing decisions and marketing activities have on your business
	**TS Filler**	This tool allows a user to take a data stream of time series data and "fill in" any gaps in the series	This tool is used primarily as a preparation step for using downstream time series-related tools and macros. Some time series tools will produce unexpected results or errors if the data stream contains gaps in the time series, e.g. you have a series of data that is supposed to contain measurements every 5 minutes, but you don't actually have measurements covering every 5 minutes.
	**TS Covariant Forecast**	The TS Covariate Forecast tool provides forecasts from an ARIMA model estimated using covariates for a user-specified number of future periods. In addition, upper and lower confidence interval bounds are provided for two different (user-specified) percentage confidence levels. For each confidence level, the expected probability that the true value will fall within the provided bounds corresponds to the confidence level percentage. In addition to the model, the covariate values for the forecast horizon must also be provided	For example, if a home appliance maker is interested in forecasting sales levels in their commercial sales division (i.e., appliances for new residential construction) then an important covariate for forecasting sales is likely to be past values of residential housing starts since appliances are among the last items placed into a new home. This tools allows for creation of forecasts from ARIMA models that include covariates
	**TS Forecast**	Provide forecasts from either an ARIMA or ETS model for a specific number of future periods.	This tool can be used for inventory management. For instance based on past history and inventory levels to help predict what your inventory levels should be in the next 3 months. The forecasts are carried out using models created using either the TS ARIMA or TS ETS tools.
	**TS Plot**	Create a number of different univariate time series plots, to aid in the understanding the time series data and determine how to develop a forecasting model.	This tool allows a number of different, commonly used time series plots to be created. One particularly interesting plot is the Time Series Decomposition plot breaks a time series into longer term trend, seasonal, and error components. This can be particularly useful in spotting changes in underling trend (say the use of a particular cell tower) in what is otherwise "noisy" data due to time of day and day of week effects.

Predictive Grouping

Icon	Tool	Description	Example
	**Append Cluster**	Appends the cluster assignments from a K-Centroids Cluster Analysis tool to a data stream containing the set of fields (with the same names, but not necessarily the same values) used to create the original cluster solution.	After you create clusters using a cluster analysis tool you can then append the cluster assignments both to the database used to create the clusters as well as to new data not used in creating the set of clusters.
	**K-Centroids Analysis**	Partition records into "K groups" around centroids by assigning cluster memberships, using K-Means, K-Medians, or Neural Gas clustering.	Common applications of this method are to create customer segments based on past purchasing behavior as a way of simplify the task of developing specialized marketing program creation (a firm can hope to develop six specialized marketing programs, but not a specialized program of each of 30,000 customers) and the creation of store groups by retailers based on store and trade area characteristics (ranging from demographics to weather patterns) for implementing merchandising, promotion, inventory, and pricing programs.
	**K-Centeroids Diagnostics**	Assess the appropriate number of clusters to specify, given the data and the selected Predictive Grouping algorithm (K-Means, K-Medians, or Neural Gas).	This tool is designed to help the user determine the "right" number of clusters to use in creating segment or groups To do this, information on the stability of the clusters (i.e., whether small changes in the

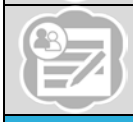
These tools are new in 8.5 8.6 9.0 9.1 9.5

			underlying data result in the creation of similar or very different clusters) and the ratio of cluster separation (how far apart the cluster are to one another) to cluster compactness (how close the members of a cluster are to one another) are provided to the user for different clustering solutions (i.e., across different possible number of clusters and methods). By comparing these measures across different clustering solutions, the user can make an informed judgment about the number of segment or groups to create. Ideally, the resulting groups should have a high level of stability and a high ratio of separation to compactness
	**K-Nearest Neighbor**	Find the selected number of nearest neighbors in the "data" stream that corresponds to each record in the "query" stream based on their Euclidean distance.	Customers that have spent various amounts of money on your product, and categorize them into “big spender”, “medium spender”, “small spender”, and “will never buy anything” categories. K-Nearest Neighbor can be used to help determine what category to put a new user before they’ve bought anything, based on what you know about them when they arrive, namely their attributes and how people with similar attributes are categorized.
	**Principal Components**	Reduce the dimensions (number of numeric fields) in a database by transforming the original set of fields into a smaller set that accounts for most of the variance (i.e., information) in the data. The new fields are called factors, or principal components.	Groups fields together by examining how a set of variables correlate or relate with one another- such as household income level and educational attainment for people living in different geographic areas. For a particular area we may have the percentage of people who fall into each of eight different educational groups and the percentage of households that fall into nine different income groups. It turns out that household income and educational attainments are highly related (correlated) with one another. Taking advantage of this, a principal components analysis may allow a user to reduce the 17 different education and income groups into one or two composite fields that are created by the analysis. These two composite fields would capture nearly all the information contained in the original 17 fields, greatly simplifying downstream analyses (such as a cluster analysis).

Connectors

Icon	Tool	Description	Example
	**Amazon S3 Download**	Read CSV, DBF and YXDB files from Amazon S3.	Get data that is stored in Amazon S3- useful for the Analytics Gallery- b/c it is hosted in Amazon
	**Amazon S3 Upload**	Write CSV, DBF and YXDB files to Amazon S3.	Put data in Amazon S3
	**Download**	Retrieve data from a specified URL, including an FTP site, for use in a data stream.	Pull data off the web. Download competitors store listings from their website
	**Google Analytics**	Bring in data from Google Analytics	A company can combine google analytics data with other sources of data.
	**HDFS Input**	Reads data from a Hadoop Distributed File System. It is able to retrieve *.csv and *.avro files. All Hadoop distributions implementing the HDFS standard are supported	The HDFS input tool allows you to access data stored in Hadoop for building data blending and advanced analytics workflows.
	**HDFS Output**	writes data to a Hadoop Distributed File System. All Hadoop distributions implementing the HDFS standard are supported.	The HDFS Output tool allows you to write the results of data blending and advanced analytics workflows to Hadoop.

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Marketo Input**	The Marketo Input Tool reads Marketo records for a specified date range. Two types of Marketo records can be retrieved: 1. **LeadRecord:** These are lead records and there will be one record for each lead. 2. **ChangeRecord:** These records track the activities for each lead. There are potentially many ChangeRecord Records for each LeadRecord The Input Tool retrieves records in batches of 1000 records, where the Marketo Append tool makes an API request for each record	Allows users to access data directly from Marketo
	**Marketo Append**	The Marketo Append tool retrieves Marketo records and appends them to the records of an incoming data stream. Two types of Marketo records can be retrieved: 1. **LeadRecord:** These are lead records and there will be one record for each lead. 2. **ActivityRecord:** These records track the activities for each lead. There can be many ActivityRecord records for each LeadRecord. Both of these record types are retrieved by specifying a LeadKey, which must be supplied by an upstream tool. More information on LeadKeys can be found in the Append Tab section under Configuration Properties.	Appends records accessed from Marketo
	**Marketo Output**	The Marketo Output tool calls to the Marketo API function: syncLead(). Data is written back to Marketo using an 'Upsert' operation. This means if a record doesn't currently exist, it will be created (see 'Inserts' below). If a record currently exists, it will be updated	Allows users to write data back into Marketo
	**Mongo DB Output**	Write data to a MongoDB database. MongoDB is a scalable, high-performance, open source NoSQL database.	Allows users to writing data back into MongDB
	**MongoDB Input**	Read and query data from a MongoDB database. MongoDB is a scalable, high-performance, open source NoSQL database.	Allows users to read data that is stored in MongoDB- can be used to enable Big Data analytics such as Ecommerce transaction analysis.
	**Salesforce Input**	Read and query data from Salesforce.com.	Pull data from SFDC Gives organizations better insight into their customers/users to plan better marketing and sales activities
	**Salesforce Output**	Write data to Salesforce.com.	Write to SFDC tables from Alteryx
	**SharePoint List Input**	Read a list from SharePoint.	Input data from SharePoint list
	**SharePoint List Output**	Write data to a list in SharePoint.	Write data stream to a SharePoint list
Address			
Icon	**Tool**	**Description**	**Example**
	**CASS**	Standardize address data to conform to the U.S. Postal Service CASS (Coding Accuracy Support System) or Canadian SOA (Statement of Accuracy).	This can be used to help in Marketing optimization by improving on the accuracy of the address and customer information – can improve geocoding and fuzzy matching by standardizing address data – CASS

These tools are new in 8.5 8.6 9.0 9.1 9.5

			Certification is required by USPS for bulk mail discounts.
	**Canada Geocoder**	Determine the coordinates (Latitude and Longitude) of an address and attach a corresponding spatial object to your data stream. Uses multiple tools to produce the most accurate answer.	Could be used for processing customer files to identify clustering of customers. Geocoding – assigning an address to a physical location is the first crucial step to any spatial analysis.
	**US Geocoder**	Determine the coordinates (Latitude and Longitude) of an address and attach a corresponding spatial object to your data stream. Uses multiple tools to produce the most accurate answer.	Could be used for processing customer files to identify clustering of customers. Geocoding – assigning an address to a physical location is the first crucial step to any spatial analysis.
	**Parse Address**	Parse a single address field into different fields for each component part such as: number, street, city, ZIP. Consider using the CASS tool for better accuracy.	Take anaddress field and break it into different pieces. Example, address in a single field/column- city, zip, street
	**Street Geocoder**	Determine the coordinates (Latitude and Longitude) of an address and attach a corresponding spatial object to your data stream. Consider using the U.S. Geocoder or Canadian Geocoder macros for better accuracy.	Find the latitude, longitude of a customer Could be used for processing customer files to identify clustering of customers. Geocoding – assigning an address to a physical location is the first crucial step to any spatial analysis.
	**US ZIP9 Coder**	Determine the coordinates (Latitude and Longitude) of a 5, 7, or 9 digit ZIP code.	Find the latitude longitude of a ZIP+4
Demographic Analysis			
Icon	**Tool**	**Description**	**Example**
	**Allocate Append**	Append demographic variables to your data stream from the installed dataset(s).	Append demographics to a trade area- example- what is the population within 10 mins of my store
	**Allocate Input**	Input geographies and demographics into a data stream from the installed dataset(s).	Demographic based on standard geographies- Example- population based on standard zip codes
	**Allocate Metainfo**	Input demographic descriptions and unabbreviated variable names ("popcy" is displayed as "population current year") from the installed dataset(s).	Returns with descriptions of demographics- converts the variable name (popcy) to human names, (population current year)
	**Allocate Report**	Create pre-formatted reports associated with Allocate data from the installed dataset(s).	Pre-built demographic reports- example- a Store trade area returns a formatted demographic report
Behavior Analysis Tools			
	Tool	**Description**	**Example**
	**Detail Fields**	Split a behavior profile set into its individual clusters and details.	Calculates distributions and penetration of customers by segment (data output)
	**Behavior Metainfo**	Input behavior cluster names, IDs and other meta info from an installed dataset.	Generates a list of all of the clusters in the segmentation system
	**Cluster Code**	Append a behavior cluster code to each record in the incoming stream.	Example- It appends a lifestyle segments to records based on a geo-demographic code (ie. block group code)

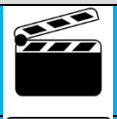
These tools are new in 8.5 8.6 9.0 9.1 9.5

Icon	Tool	Description	Example
	**Compare Behavior**	Compare two behavior profile sets to output a variety of measures such as market potential index, penetration, etc.	Calculate correlations and market potential between two profiles- Market potential for new customers in a new market that fit the same profiles of current customers
	**Create Profile**	Create behavior profiles from cluster information in an incoming data stream.	Example- it can show the distribution of customers across the clusters in the segmentation
	**Report Comparison**	Generate a comparison report from two behavior profile sets for output via the Render tool.	Shows market potential towards the likelihood of certain purchasing behaviors at the segment level
	**Report Detail**	Generate a detailed report from a behavior profile set for output via the Render tool.	Generates a report of distributions and penetration of customers by segment
	**Report Rank**	Generate a rank report from a set of behavior profiles for output via the Render tool.	Use to rank geographies based on market potential of prospective customers
	**Profile Input**	Input a behavior profile set from an installed dataset or external file.	Used in conjunction or after the Write Behavior Profile set- Provides the ability to open the previously saved profile sets and use for analysis
	**Profile Output**	Output a profile set (*.scd file) from behavior profile sets in an incoming data stream. Generally only used when using the standalone Solocast desktop tool.	Taking a customer file that has a segment appended to it and saving it into an Alteryx format so that it can be used repeatedly
Calgary (List Count Retrieval Engine)			
Icon	**Tool**	**Description**	**Example**
	**Cross Count**	Find the counts of predefined sets of values that occur in a Calgary database file.	Very fast, find a count of how many times a value occurs in a file- this can be filtered or based on a user defined condition
	**Cross Count Append**	Find the counts of sets of values (from the incoming data stream) that occur in a Calgary database file.	Very fast, find a count based on a filter that is gone into the tool from previous process. Example- number of people by segmentation group in a trade area(the advantage is you are getting a count, rather than extracting the data and getting the count yourself)
	**Calgary Input**	Input data from the Calgary database file with a query.	Very Fast input data from an indexed file. This can be filtered or based on a user defined condition
	**Calgary Join**	Query a Calgary database dynamically based on values from an incoming data stream.	Very fast extraction data from a large index file, based on a condition passed to it from inside Alteryx. Example- extract all businesses with employees greater than 1000 (where 1000 is from an Alteryx calculation above it)
	**Calgary Loader**	Create a highly indexed and compressed Calgary database which allows for extremely fast queries.	Create a highly indexed and compressed file for use with any of the Calgary tools (to enable the speed)
Developer			
Icon	**Tool**	**Description**	**Example**

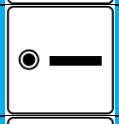
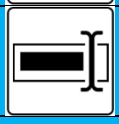
These tools are new in 8.5 8.6 9.0 9.1 9.5

	**API Output**	Return the results of a data stream directly to an API callback function. For use with custom application development.	Have an Alteryx process being called from the inside of another program. Example- someone wants to have their own Application interface, but use Alteryx for the processing. (ie. Mortgage calculator)
	**Base64 Encoder**	The Base 64 Encoder macro issues a base 64 encode string	Useful where there is a need to encode binary data that needs to be stored and transferred over media that is designed to deal with textual data. This is to ensure that the data remains intact without modification during transport. (http://en.wikipedia.org/wiki/Base64) This Macro is used in our Social Media tools
	**Block until Done**	Stop downstream processing until the very last record has arrived, to ensure that only a single output stream processes records at one time. Or, ensure that the reading of a file will be closed before overwriting is attempted.	Used when you want to make sure part of the process is finished before the next part of the process begins
	**Detour**	Bypass a set of tools. Must end in an Output or Detour End tool. Generally used for authoring an Analytic App or Macro.	Only used in apps where you are giving an option to an end user to this or that.
	**Detour End**	Ends a section of tools bypassed by a Detour. Generally used for authoring an Analytic App or Macro.	Connect with use of Detour
	**Dynamic Input**	Read from input files or databases at runtime using an incoming data stream to dynamically choose the data. Allows for dynamically generated queries.	You have a process in Alteryx that can pull a value/extract from a Database. Example, you have a process in Alteryx that says which zip codes have the highest value. Then that extracts data from another source based on that zip code – decreases overhead of reading in the entire database
	**Dynamic Rename**	Dynamically (using data from an incoming stream) rename fields. Useful when applying custom parsing to text files.	If you bring in a csv file, where all of the column names have underscores, this could replace all of the underscores to spaces. It will not change the data, but will change the column names.
	**Dynamic Replace**	Replace data values in a series of fields (using a dynamically specified condition) with expressions or values from an incoming stream.	Say you have a hundred different income fields and instead of the actual value in each field, you want to represent the number with a code of A, B, C, D, etc. that represents a range. The Dynamic Replace tool can easily perform this task.
	**Dynamic Select**	Select or de-select fields by field type or an expression.	Select the first 10 columns of a file regardless of what they contain
	**Field Info**	Output the schema (field types and names, etc.) of a data stream.	Gives users the types of the data, that the column contains(description of the data)
	**JSON Parse**	The JSON Parse tool separates Java Script Object Notation text into a table schema for the purpose of downstream processing. It can be built back up into usable JSON format by feeding the output into the JSON Build tool.	This tool is used in our Social Media tools where we are pulling data from web API's that return JSON via the Download tool. The JSON is parsed before it is presented to the user into a nicely formatted table.
	**Message**	Write log messages to the Output Window. Generally used in authoring macros.	Gives you extra messaging information in the output. The number of columns in a certain table, the number of rows, or errors.
	**R**	Execute an R language script and link incoming and outgoing data from Alteryx to R, an open-source tool used for statistical and predictive analysis.	This free open-source programming language has 1000's of algorithms available, or you can write your own algorithm in R and bring it into Alteryx
	**Run Command**	Run external programs as part of an Alteryx process.	Call other programs from inside Alteryx- example, curl.exe was widely used to scape data, before Alteryx had the download tool

These tools are new in 8.5 8.6 9.0 9.1 9.5

Icon	Tool	Description	Example
	**Test**	Test assumptions in a data stream.	Check conditions inside modules and give errors if required- example- if you want to make sure all of the records are joined before you start the process, the error/test would stop the module
Social Media			
Icon	**Tool**	**Description**	**Example**
	**Foursquare**	Search Foursquare Venues by a location with an option of filtering by a search term.	A company can analyze location data of where users are checking in.
	**Gnip Input**	Bring in data collected from twitter, youtube, facebook in Gnip and analyze it in Alteryx	A company can analyze what people are saying around their company over that last week.
	**Twitter Search**	Search tweets of the last 7 days by given search terms with location and user relationship as optional properties.	A company can analyze what people are saying around their company over that last week.
	**Datasift**	Bring in data from Datasift- twitter, facebook, tumblr, youtube, Wikipedia, and much more- http://datasift.com/platform/datasources/	A company can analyze what people are saying around their company from various sources of media.
Laboratory			
Icon	**Tool**	**Description**	**Example**
	**Blob Convert**	The Blob Convert tool will take different data types and either convert them to a Binary Large Object (Blob) or take a Blob and convert it to a different data type.	Convert a PNG, GIF or JPG Blob to Report Snippet. Allows the Reporting Tools to recognize the incoming images as report snippets for building reports.
	**Blob Input**	The Blob input tool will read a Binary Large Object such as an image or media file, by browsing directly to a file or passing a list of files to read.	Read a list of documents from disk either as blobs or encoded strings.
	**Blob Output**	The Blob Output tool writes out each record into its own file	Write a list of documents to disk either as blobs or encoded strings.
	**JSON Build**	The JSON Build tool takes the table schema of the JSON Parse tool and builds it back into properly formatted Java Script Object Notation.	After parsing JSON data, send it through the Formula tool to modify it as needed and then use the JSON Build tool to output properly formatted JSON.
	**Make Columns**	The Make Columns tool takes rows of data and arranges them by wrapping records into multiple columns. The user can specify how many columns to create and whether they want records to layout horizontally or vertically.	This tool is useful for reporting or display purposes where you want to layout records to fit nicely within a table. Arrange a table of 10 records into a table of 5 records spread across 2 columns.
	**Throttle**	The Throttle tool slows down the speed of the downstream tool by limiting the number of records that are passed through the Throttle tool.	This is useful for slowing down requests sent per minute when there are limits to how many records can be sent in.
Interface			
Icon	**Tool**	**Description**	**Example**
	**Action**	The Action tool updates the configuration of a module with values provided by interface questions, when run as an app or macro.	
	**Check Box**	The Check Box tool will display a check box option to the end user in an app or macro. The resulting value, True (checked) or False (unchecked), is passed to downstream tools.	

These tools are new in 8.5 8.6 9.0 9.1 9.5

	**Condition**	The Condition tool tests for the presence of user selections. The state is either true or false.	
	**Control Parameter**	Control Parameter tool is the input for each iteration in the Batch Macro. The Control Parameter input will appear as a ¿ below its input arrow on the macro tool icon. For each record coming into the Control Parameter input, the entire macro will be re-configured and run beginning to end.	
	**Date**	The Date tool will display a calendar in app or macro for the end user to specify a date value. The resulting date value is passed through downstream tools	
	**Drop Down**	The Drop Down tool will display a single selection list in an app or macro to an end user. The single value selection is then passed to downstream tools.	
	**Error Message**	The Error Message tool will throw an Error message to the end user of an app or macro. Once the error message is thrown, all downstream processing stops.	
	**File Browse**	The File Browse tool will display a file browse control in an app. This tool can be used to read an input or write an output. The value of the file path specified by the user is passed to downstream tools.	
	**Folder Browse**	The Folder Browse tool displays a folder browse control in an app or macro. The directory path specified by the user is passed to downstream tools.	
	**List Box**	List Box tool will display a multi-selection check box list in an app or macro. The selections specified by the user are passed as values to downstream tools.	
	**Macro Input**	The Macro Input tool is used to display input parameters (arrows, source, and interface) for a macro.	
	**Macro Output**	The Macro Output tool is used to display output arrows on a macro tool.	
	**Map**	The Map tool will display an interactive map for the user to draw or select map objects in an app or macro. The map selection made by the user is passed to downstream tools.	
	**Numeric Up Down**	The Numeric Up Down tool is used to display a numeric control in an app or macro. The value specified by the user is passed to downstream tools.	
	**Radio Button**	The Radio Button tool will display a mutually exclusive option in an app or macro. The resulting value, True (ticked) or False (not ticked), is passed to downstream tools.	
	**Text Box**	The Text Box tool will display a free form text box in an app or macro. The user entered value will be passed to downstream tools.	
	**Tree**	The Tree tool will display an organized, hierarchal data structure in an app or macro. The selections made by the user are passed as values to downstream tools. The values returned from trees are separated by a new line character (\n).	

These tools are new in 8.5 8.6 9.0 9.1 9.5

In-Database Tools			
Icon	**Tool**	**Description**	**Example**
	**Browse Data In-DB**	Review your data at any point in an In-DB workflow. Note: Each In-DB Browse triggers a database query and can impact performance.	Use the Browse Data In-DB tool as you build an In-DB workflow to ensure data is passing through the way you intend it to. It allows you to view results at any point in the In-DB workflow
	**Connect In-DB**	Establish a database connection for an In-DB workflow	If you want to connect directly to Oracle or SQL Server.
	**Filter In-DB**	Filter In-DB records with a Basic filter or with a Custom expression using the database's native language (e.g., SQL).	If you are looking at a dataset of customers, you may want to only keep those customer records that have greater than 10 transactions or a certain level of sales and want this process to run in a database.
	**Formula In-DB**	Create or update fields in an In-DB data stream with an expression using the database's native language (e.g., SQL).	For instance if there is missing or NULL value, you can use the formula tool to replace that NULL value with a zero and run this process in a database. Basically all of the formulas you can do in excel you can do inside Alteryx formula tool – so many more!
	**Join In-DB**	Combine two In-DB data streams based on common fields by performing an inner or outer join.	For instance this can be used to join Customer profile data as well as transactional data, and join the two data sources based on a unique customer ID and run this process in a database.
	**Sample In-DB**	Limit the In-DB data stream to a number or percentage of records.	Choosing the first 10 records for each region of previously sorted data so that you end up with the top 10 stores in each region and run this process in a database.
	**Select In-DB**	Select, deselect, reorder, and rename fields in an In-DB workflow.	If your workflow only requires 5 fields out of 50 that are read in from the file/database, you can deselect all but the 5 required fields to speed up processing downstream in a database.
	**Data Stream In**	Bring data from a standard workflow into an In-DB workflow.	If you already have an establish Alteryx Workflow and you want to bring data into a database.
	**Data Stream Out**	Stream data from an In-DB workflow to a standard workflow, with an option to sort the records.	If you already have an establish Alteryx Workflow and you want to bring data from a database into the workflow.
	**Summarize In-DB**	Summarize In-DB data by grouping, summing, counting, counting distinct fields, and more. The output contains only the result of the calculation(s).	You could determine how many customers you have in the state of NY and how much they have spent in total or an average per transaction while all of this process takes place in the database.
	**Union In-DB**	Combine two or more In-DB data streams with similar structures based on field names or positions. In the output, each column will contain the data from each input.	Transaction data stored in different files for different time periods, such as a sales data file for March and a separate one for April, can be combined into one data stream for further processing inside the database.
	**Write In-DB**	Use an In-DB data stream to create or update a table directly in the database.	If you need to update and save the results of the in-database blending into the database.