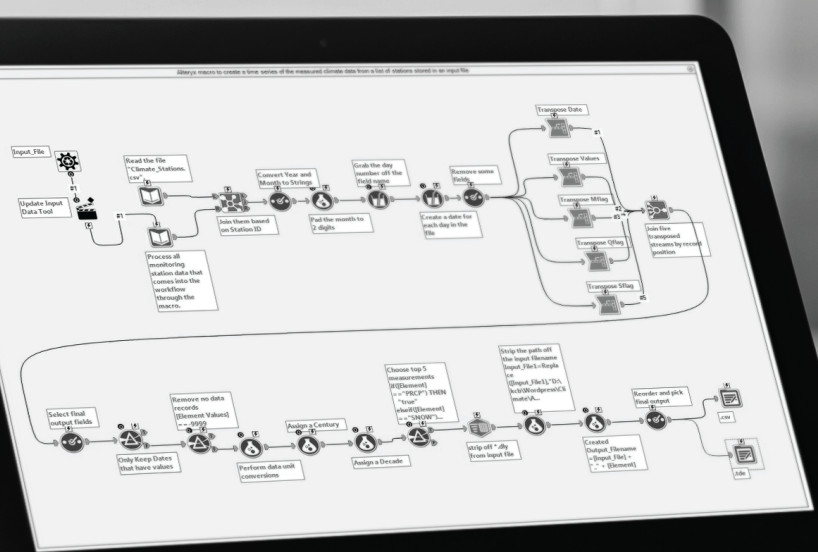
USEREADY

A PRACTITIONER'S GUIDE TO ALTERYX

PREVIEW COPY



Alteryx 11

A Practitioner's Guide to Alteryx®

A Practitioner's Guide to Alteryx®

Published by
USEReady
805 Third Avenue, 11th floor, New York, NY 10022
www.useready.com

ISBN: 978-0-692-44795-6
Manufactured in the United States of America

About USEReady

USEReady's mission is to help organizations succeed by fast tracking their business performance. We relentlessly strive towards value-driven innovation and digital transformation using advanced business intelligence, data management, and infrastructure management services. Our global delivery orientation is enriched with winning strategies, rich industrial and technology expertise and customer fanaticism. Our proven track record of success in versatile domains such as capital markets, insurance, healthcare, pharma, retail, media etc. is a testimony to our performance.

Visit www.useready.com or e-mail contact@useready.com

About Alteryx, Inc.

Alteryx, Inc. is the leader in data blending and advanced analytics software. Alteryx Analytics provides analysts with an intuitive workflow for data blending and advanced analytics that leads to deeper insights in hours, not weeks, which is typical of traditional approaches. Analysts love the Alteryx Analytics Platform because they can deliver deeper insights by seamlessly blending internal, third party and cloud data; and then analyze it using spatial and predictive drag-and-drop tools. This is all done in a single workflow, with no programming required. More than 1,000 customers and thousands of data analysts worldwide rely on Alteryx daily.

Visit www.alteryx.com or call 1-888-836-4274.

Acknowledgements

When we started writing this edition of the book, we decided to go with a team of authors instead of a single author. The team went through a pretty challenging process of reading, updating, and reviews to get the book where it is today.

I would like to express my gratitude to the team of authors and other colleagues who helped in making this book a reality. Without them, the book truly would not be what it is today.

Uday Hegde: For being the driving force and a constant motivator.

Anand Vadul: For sharing the experiences and challenges faced in the last edition and being the technical guiding force.

Prashant Singh, Suman Joshi, and Vishnudas PN: For helping with content collation, editing, marketing, and an excellent cover design.

Subrat Das

Letter from the CEO

As we are publishing this third edition, Alteryx is a public company. Alteryx has gained a wide spread adoption across industries and geographies. We are glad that we partnered with this successful product early on. We see an increased adoption through the orders that we receive for this book. This edition incorporates many feedback and suggestions received from our readers.

As suggested by many of you, we have incorporated additional content by adding five new chapters and several hands-on exercises. Many newly certified Alteryx practitioners at USEReady have contributed to this book. Their hard work and dedication has resulted in this book that is now reaching 1000 pages in content. A total of 22 practitioners have revised this book with upgraded product features, exercises and new chapters.

The team has gone great heights to ensure the content is refined to the most recent version of Alteryx and exercises are useful to a fellow practitioner.

We are grateful to our readers of the previous editions and their constructive feedback has helped us improve this edition. We hope our efforts are well worth it and you are going to find this book useful.

Uday Hegde
Chief Executive Officer
USEReady

Foreword

As Vice President of product management at Alteryx, Inc. working closely with our product team, our customers and our partners, a training manual from USEReady is testament to the growing demand for easy-to-use data blending and advanced analytics solutions. USEReady knows the analytics industry and has created this book as both an independent guide and as a classroom aid to help its customers and others not only quickly learn Alteryx products, but more importantly grow in their ability to help gather deeper insights from their data.

The book has created a great launching point for beginners who want a manual, in addition to the real-world use cases, so you can easily learn how best to use Alteryx Designer. By the time, you finish working through this manual, someone who has never opened the Alteryx Designer before, will be able to create workflows, design reports, develop applications, and write macros to solve any of their data needs.

These are exciting times as the Alteryx community continues to grow globally, and we continue to witness an unprecedented demand for data analytics with actionable information. The Alteryx Designer and materials like this manual help the community of self-service data analysts make the most of their data.

We appreciate the partnership with companies like USEReady that know how to help these self-service analysts and reduce the time to insight with Alteryx.

Laura Sellers
Vice President, Product Management of Alteryx

Preface

Every day we are faced with options, questions, and choices. These decisions, as we all know, are much easier to make when we are well informed. Let's say that we want to eat. We literally have an entire world of possibilities, given the proper resources. However, practically, there are real limitations. Are we at home without transportation? Are we backpacking in the mountains? Are we in the middle of Times Square? Do we have food restrictions for health reasons? Do we have $5 or $5000? What are we in the mood for?

This task that we all solve day in and day out depends on a considerable amount of information that we know about our world, and often take for granted. This information is all based on data about our world.

What Is Data?

Data is stored information. It comes in various forms ranging from the number and types of items on our desk to the total mass of the universe to the contents of this book to the information in digital files and systems, which will be our focus.

What Does Data Do?

Data does nothing. It simply exists. It is what we do with data that is important. When we look at data we interpret it to create meaningful information, which gives us the ability to make better-informed decisions.

How Do We Consume Data?

Data can be consumed in many forms. We can look at all of the raw data and read every piece individually. We can use aggregation methods to create summary data so that we can

easily see high-level trends. We can visualize the data because a picture truly is worth a thousand words. Since we often do not want to look directly at the original data source and read each individual piece of data, we need to perform data preparation.

What Is Data Preparation?

Data Preparation is the process by which raw data is converted into a clean, usable source for later consumption.

The three core components of data preparation are data retrieval, data manipulation, and data export. In more traditional analytic terms, data preparation refers to the extract-transform-load process referred to as ETL. However, in order to ease communication, we are going to avoid these technical terms and discuss the aspects of the processes as follows:

Data retrieval refers to the process of going to a data source, asking for data, and returning the desired data.

Data manipulation refers to anything we decide to do to the data between the time we retrieve it and the time we export it.

Data export refers to what we do with data after we have extracted and manipulated it even if we haven't finished transforming it.

What is Data Manipulation?

Data retrieval and Data export are fairly straightforward; respectively, they can be likened to drawing water from a well and putting an ice cube in someone's drink. However, data manipulation is that tricky process of running the water through the pipes into our house, then filling the ice cube tray, then putting the tray in the freezer, and letting the water have enough time to freeze so that we have ice to consume. Going forward, we

will be using an allegory to a river to explain the entire process of the data preparation and specifically the data manipulation portion.

Data manipulation can come in many forms which typically fall into three buckets as follows:

Combination

One of the most common problems with data is that it comes from multiple sources. It is generally possible to perform the analysis separately, or through a significant amount of manual effort, but these methods often leave something to be desired or are too slow for effective use. In order to solve this problem, we will be designing data streams that come together.

If we think about data streams as actual rivers, original data locations can be thought of as glacial streams, smaller rivers, or lakes. Bringing data together is like the tributaries that bring these different water sources together to form a river. Along the course of this river, way we can perform calculations.

Calculation

If the data is to be used, it is generally advisable to have as much data pre-calculated as possible. One reason for this is that it allows an organization to create a standardized formula for everyone's use. Another is that when we can run calculations before data is provided to a front-end user or system, the consumer will experience a much faster process.

If the data is to be used in a report, then the calculations are often fundamental aspects of that report.

Returning to the river analogy, we can think of calculations as hydroelectric dams along a river, we are using the resources that already exist in order to generate something new.

We may change the landscape because we are changing the flow of water, and we are also slowing down the river (introducing calculations will slow down the data preparation process).

We also have the ability to transform the data stream into a more useable format.

Transformation

It is often the case that data is not in the format that we need. We may have been given access to a database that has data stored in a very machine readable format, and we need to pivot the table to make it human readable, or we may have been working with an Excel file which has data extremely normalized that makes it hard to use in a front-end system. Either way, we need to transform the structure of the data so that it can be effectively consumed. In thinking about the river, we can imagine this as the process of cutting a channel into the riverbed so that the river is deep enough to move barges up and down. In doing this, we are fundamentally changing the structure of the river in order to make it more useable.

To the reader

In the following chapters, we are going to cover many topics, but the format of the chapters will all be the same.

You will assume the role of a new consultant at a company that works with Alteryx. We introduce a business scenario, discuss the tools that we will use to solve the problem(s), walk through the initial problem(s), and then provide you a self-guided exercise. We conclude this book with a capstone assignment in *NYC*.

The exercises will use data that can be download from https://resources.useready.com/publications/a-practitioners-guide-to-alteryx-alteryx-version-11/ by following the instructions on the website to unpackage the file.

Additional data will also be needed to install the *US 2010 Census SF1* and *USGS North America Map* packages from http://downloads.alteryx.com/data.html which we will start using in *Cultural Musings*. We will also be using the Solocast Datasets in the *Statistics in Alteryx* section.

Let us know what you think by emailing us at AlteryxBook@USEReady.com and we will try to incorporate reader requests going forward.

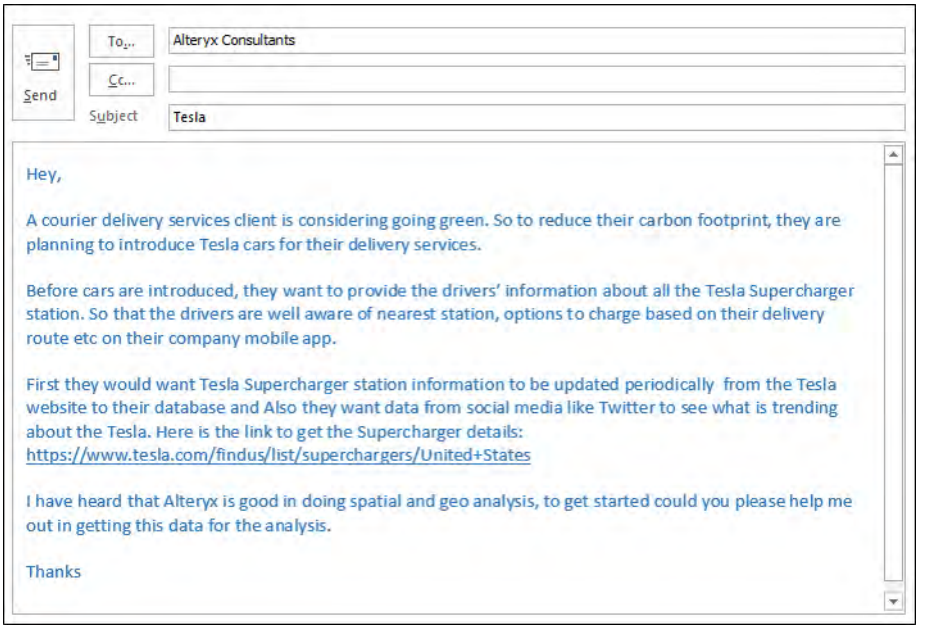
If interested in Alteryx training sessions or Alteryx consulting, visit http://www.useready.com/.

Best of Luck,

USEReady

Table of Contents

CHAPTER 12
Green On The Go


Send
To...
Alteryx Consultants
Cc...
Subject
Tesla
Hey,
A courier delivery services client is considering going green. So to reduce their carbon footprint, they are planning to introduce Tesla cars for their delivery services.
Before cars are introduced, they want to provide the drivers' information about all the Tesla Supercharger station. So that the drivers are well aware of nearest station, options to charge based on their delivery route etc on their company mobile app.
First they would want Tesla Supercharger station information to be updated periodically from the Tesla website to their database and Also they want data from social media like Twitter to see what is trending about the Tesla. Here is the link to get the Supercharger details:
https://www.tesla.com/findus/list/superchargers/United+States
I have heard that Alteryx is good in doing spatial and geo analysis, to get started could you please help me out in getting this data for the analysis.
Thanks

12.1 Tools & Concepts

Tools	Concepts
Download	Retrieve data from cloud
Amazon-S3 Download	Retrieve data from internet/intranet environment
Amazon-S3 Upload	Push data to the cloud
Google Analytics	Push data to the internet /intranet environment
Foursquare Search	
Marketo Append	
Marketo Input	
Marketo Output	
MongoDB Input	
MongoDB Output	
Salesforce Input	
Salesforce Output	
SharePoint List Input	
SharePoint List Output	
Twitter Search	

12.2 Download

Figure 12.1– Download

The *Download* tool will retrieve data from a specified URL to be used in downstream processing or to be saved to a file.

Group	Input	Output
Connectors	URL	Data field or output to a file

The *Download* tool can also download or upload data via FTP and SFTP.

Note: To avoid connection delays when using the Download tool, ensure that the "Automatically detect settings" option is deselected within your account's internet properties (Control Panel > Internet Options > Connections > LAN Settings).

An *Action* tool can be connected to the *Lightning Bolt Anchor* to modify how this tool works in apps and macros.*:*

Properties Window:

There are four tabs with configurable options on the *Download* tool: *Basic*, *Headers*, *Payload* and *Connection*. Only the *Basic* tab is required for proper configuration.

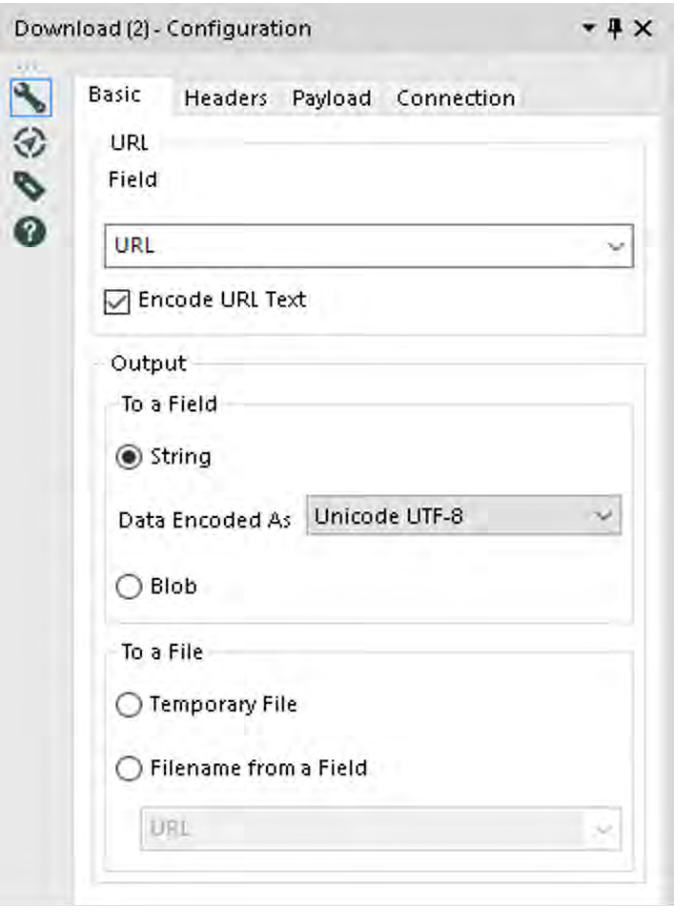


Figure 12.2 – Download Configuration

- *URL Field*: Specify the field from the incoming data stream that contains the URL to pull data from.

- *Encode URL Text*: When checked, the specified URL will be encoded as needed where unsafe ASCII characters are converted into a format that can be transmitted over the internet. An example of this would be the substitution of %20 for a space.

Output: Specify how the returned data should be formatted. The data can be returned in a data field or output to a file.

- *To a Field*: Downloaded content is returned in the data stream as a data field. The downloaded contents will be in a single field called "DownloadData". You will likely have to parse this data using downstream tools such as the Text to Columns, RegEx, or Formula tools.
    - String: Data is returned as a new wide string type field. A wide string supports Unicode characters.
    - Blob: Data is returned as a new blob type field. Blob is also known as Binary large objects. Image files are usually stored in this format. To use the image, configure a Report Image tool downstream and specify the Image or Blob field.
- *To a File:*
    - *To a Temporary File*: Data is output to a temporary file and will be located in the user's temporary directory. For more on temp file handling in Alteryx, review the Temp Files page.
    - *Filename from a Field*: Data is output to a specific file where the file specification is in an incoming field. Use the dropdown to select the field that contains the file name to output to.

The *headers tab* allows you to modify the HTTP headers sent with the web request.

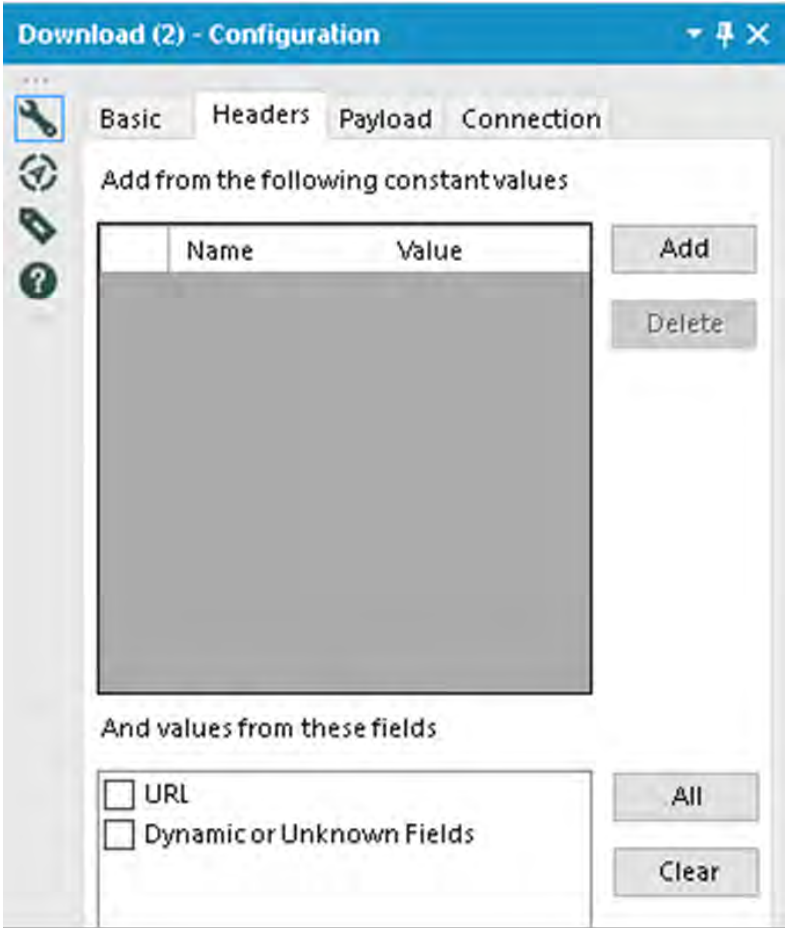


Figure 12.3–Download Configuration

- *Add from the following constant values*: Allows adding fixed header values.
- *And values from these fields*: Takes values from the record data and creates header values.

The *payload tab* allows you to set the HTTP Action you would like to perform and optionally build the Query String or Body for the web request.

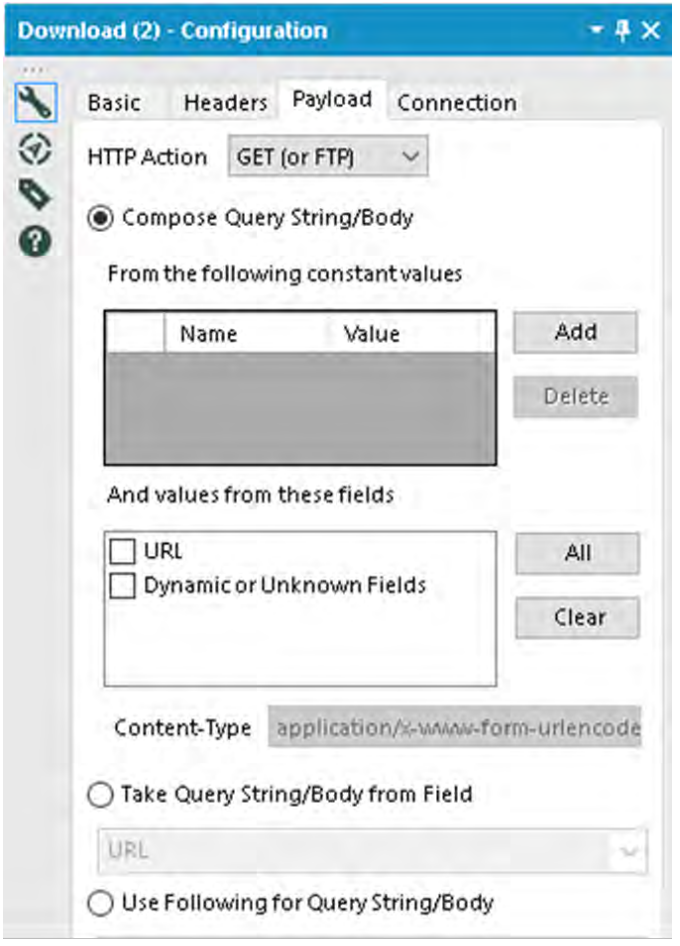


Figure 12.4–Download Configuration

- HTTP Action: Select the HTTP Action for the web request. Choices include:
    - GET (or FTP): Perform a GET request or download a file from an FTP or SFTP site. This option can be used without entering any other options on this tab and will download the selected URL.

- POST: Performs a POST request to the selected URL. Typically when using this option you would specify a POST body using the below options.
- PUT: Performs a PUT request. This option only allows you to take the query body from a blob field via the Blob Input tool. Typically used to upload a file to the remote server.
- DELETE: Performs a DELETE request. Typically used to ask the remote server to delete a specified resource.
- HEAD: Performs a HEAD request. This asks the server to return the header data, but not the body data.
- Custom: Allows you to enter a custom verb in a text box. Everything else about this request behaves the same as a POST. The POST verb is replaced by the custom verb just before the request is made.

These options will only work if the remote server (that you are sending the requests to) supports a particular verb. Check the API documentation of the URL you are using to see what requests are supported.

Choose from the following Query String / Body Options:

Compose Query String/Body:

From the Following Constant Values: Adds constant name value pairs to the query string/body

And values from these fields: Takes name value pairs from the incoming data record. The field name is the name used in the Query String/Body

Content-Type: Controls how the name value pairs are encoded.

Application/x-www-form-urlencoded: e.g. Name1=Value1&Name2=Value2&Name3=Value3. Unsafe ASCII characters are automatically encoded, so make sure your data is not already encoded.

Multipart/form-data: Only available when using HTTP Actions POST and Custom.

Take Query String/Body from Field: Select the field in the input data which contains the Query String/Body data. Blob fields can be selected when using HTTP Actions POST, PUT and Custom.

Use Following for Query String/Body: Text box allowing you to manually type in the Query String or Body contents.

All text data is UTF-8 encoded before being sent to the remote web server.

Connection tab:

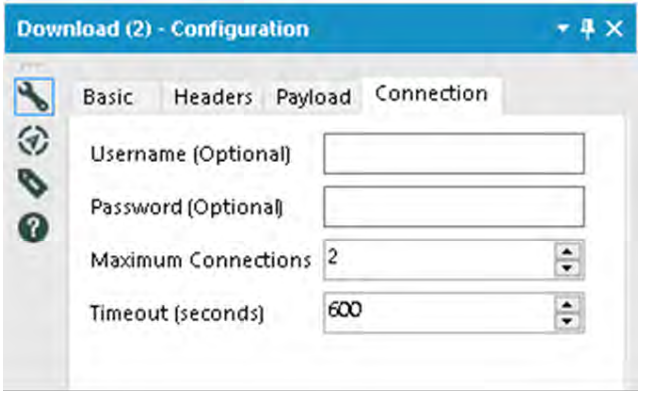


Figure 12.5–Download Configuration

- *Username*: Enter the user name if required by the URL specification above.
- *Password*: Enter the password if required by the URL specification above.
- *Maximum Connections*: Specify the maximum number of simultaneous transfers for the Download tool to perform. Transfers are only done in parallel when there are multiple input records sent to the Download tool. Multiple Download tools operate independently but do not typically function at the same time. For new Download tools added to a workflow, the default number of connections is 2 and the maximum number of connections is 32.

  Increasing the number of connections may reduce the total time taken to complete all transfers, but please use caution not to set the number too high as it could overload the server being used. It is possible the server could quit responding, report errors, or even refuse connections if it believes you are misusing it. This is particularly important when accessing a public site that is not under your control. Most web browsers will do as many as 6 simultaneous transfers, but these would typically be relatively small transfers as a part of a web page. For a server inside your own business where you have more control over how it is configured, using a higher number of connections may be okay.

  Additionally, because the Download tool will send records downstream as transfers complete, it will likely result in a change in the order of records as they pass

through the tool. If the order matters to your workflow, make sure to sort the results or limit the number of connections to 1. Finally, please note that empty URLs will be processed ahead of those that require an actual transfer.

- Timeout (seconds): Specify the number of seconds to wait before reporting a timeout due to an unresponsive connection. Select a number from 0 (never timeout) to 10,000.

12.3 Twitter Search

Figure 12.6 – Twitter Search

The *Twitter Search* tool allows you to search tweets by given search terms, with the location as an optional property. The search will only retrieve tweets from the previous seven days.

Group	Input	Output
Connectors	None	See Output data and Output Summary

Note: Before you can you use this tool, you must register an application with Twitter. Log in to your Twitter account at https://apps.twitter.com, click "Create a new application", and complete the form (a placeholder website may be used and there is no need for a Callback URL). Once you have submitted the application form, you will be provided with a Consumer Key and Consumer Secret that you can use to configure the tool.

The search will only retrieve tweets from the previous seven days.

Output Data: The selected data from your query.

Output Summary: Summary information from your query.

Properties Window:

Configuration tab

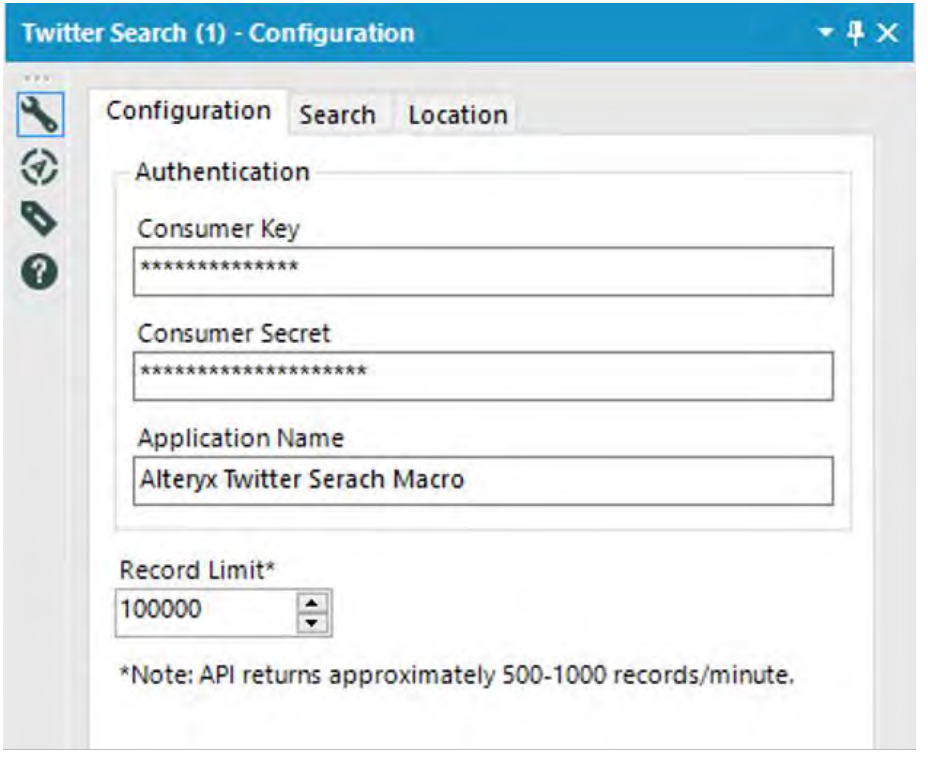


Figure-12.7 – Twitter Search Configuration

- *Authentication:* Enter your credentials.

- *Record Limit:* Specify the number of records to request, up to a maximum of 100,000 records.

Search tab with two options – Standard search and Custom Search

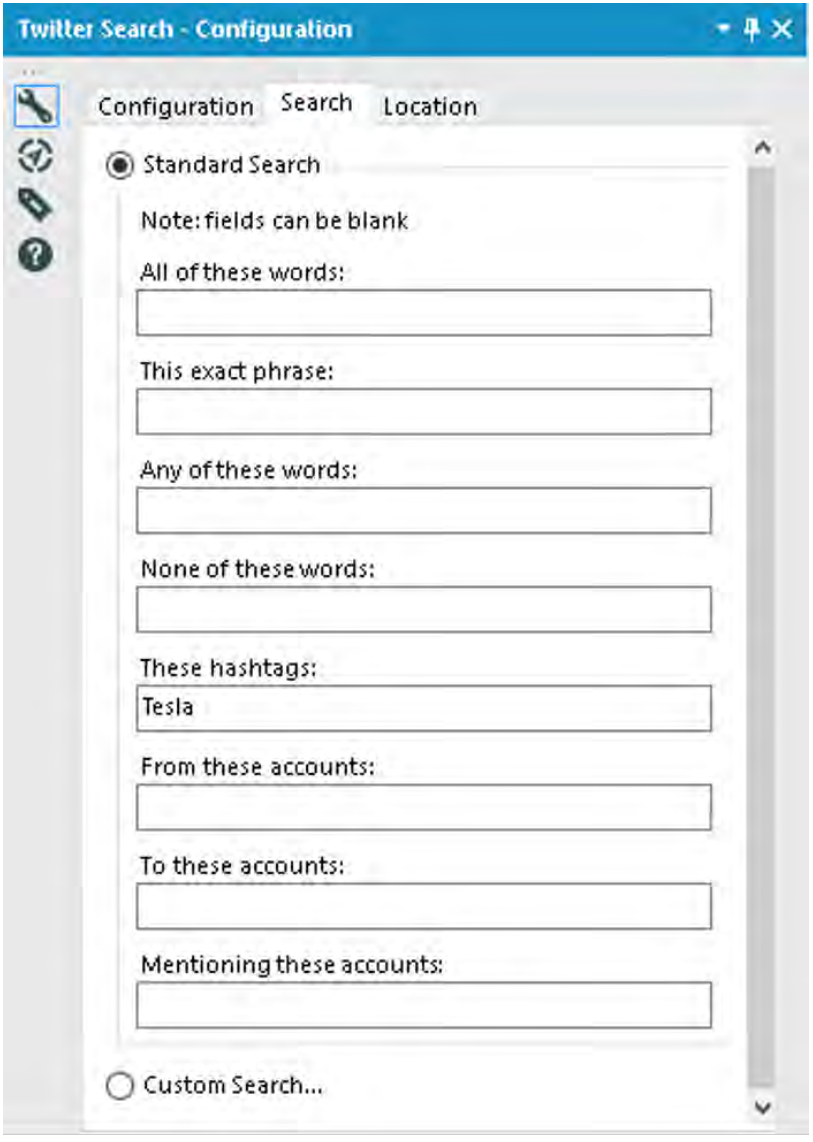


Figure 12.8–Twitter Search Configuration

If you choose *Standard Search*, you'll see eight text boxes in which to enter search criteria. You can enter criteria into one or all text boxes. The tool takes all the criteria you enter and configures it into one combined query.

- *All of these words:* Return Tweets containing all of the words (in any order) in the body of the Tweet.
- *This exact phrase:* Return Tweets containing this exact phrase in this exact order. This is one of the only text boxes in the Standard Search that allows punctuation (e.g., if the exact phrase you are searching for contains a comma or quote, do include the comma or quote in the search).
- *Any of these words:* Return Tweets containing at least one of these words.
- *None of these words:* Exclude Tweets containing any of these words.
- *These hashtags:* Return Tweets containing at least one of these hashtags.
- *From these accounts:* Return Tweets from any of these accounts.
- *To these accounts:* Return Tweets set to any of these accounts.
- *Mentioning these accounts:* Return Tweets that mention any of these accounts.

The *Custom Search* option allows users to do a more advanced search than the Standard Search by utilizing the proper Twitter API syntax. You can learn about proper Twitter API syntax via these links: https://dev.twitter.com/docs/using-search.

Location tab

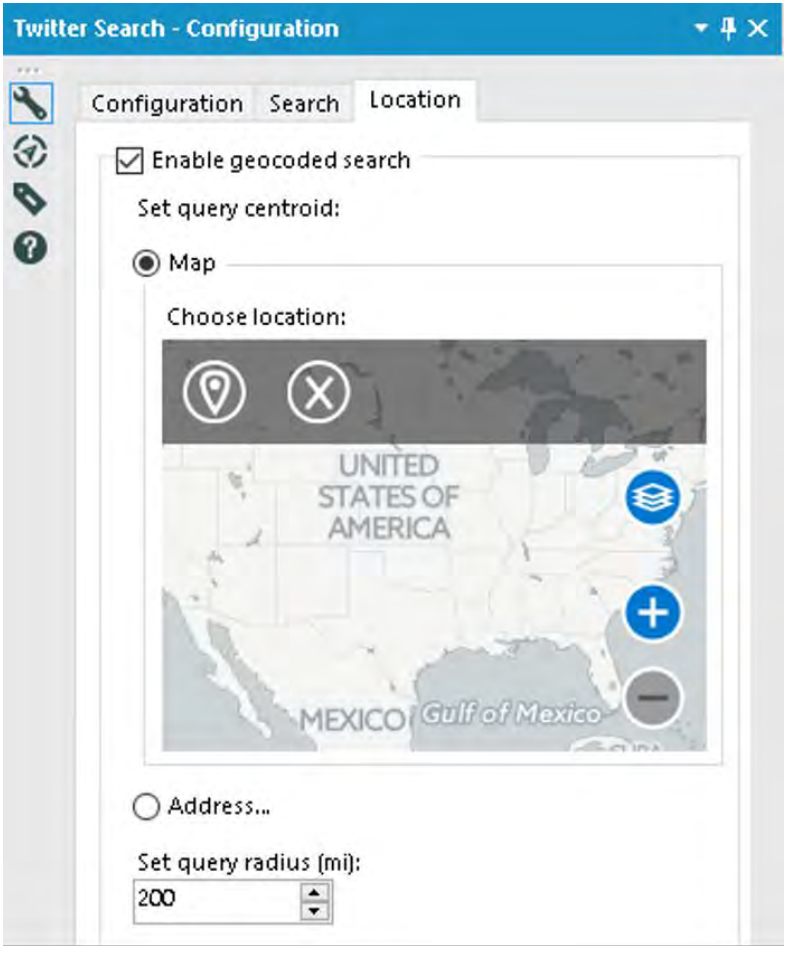


Figure 12.9–Twitter Search Configuration

- *Enable geocoded search:* Returns tweets by users located within a radius of a given latitude/longitude.
- *Set query centroid:* Specifies the latitude and longitude coordinates for the center of the search radius.

- *Map:* The geo-location of the point placed by the user on the map will be used to specify the latitude/longitude coordinates.
- *Address:* The geo-location of the address submitted by the user will be used to specify the latitude/longitude coordinates.
- *Set query radius (mi):* Sets the size of the search radius for the geo search. Defaults to five, with a max size of 1000.

12.4 Amazon S3 Download

Figure 12.10 – Amazon S3 Download

The *Amazon S3 Download* tool will retrieve data stored in the cloud where it is hosted by Amazon Simple Storage Service (Amazon S3).

Group	Input	Output
Connectors	None	Any data Stream

Output Data: Data stream containing records from cloud where it is hosted by Amazon Simple Storage Service.

Read CSV, BDF and YXDB files from Amazon S3.

Properties Window:

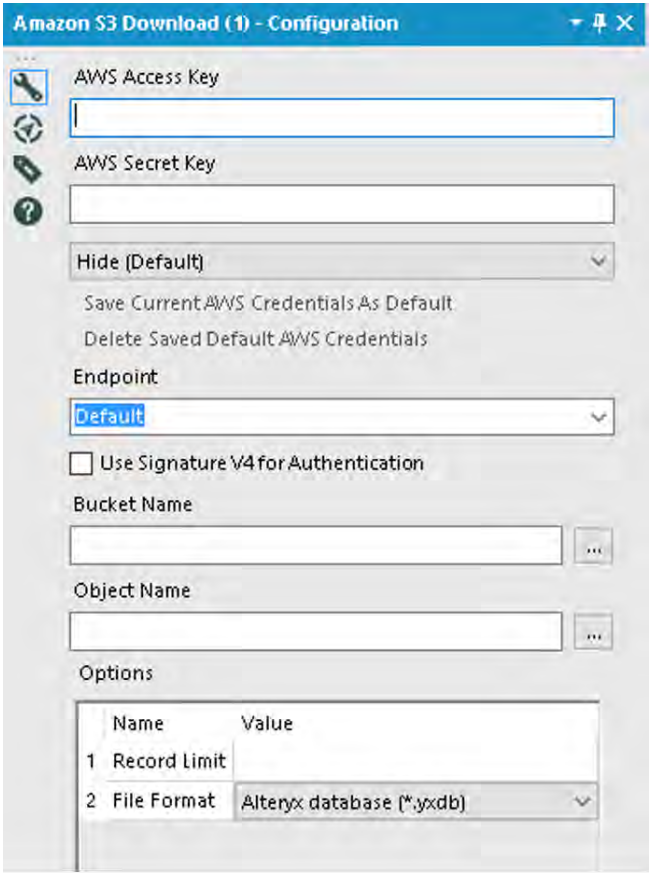


Figure 12.11 – Amazon S3 Download Configuration

- ***AWS Access Key***: Specify the Amazon Web Services Access Key to use to download data.
- ***AWS Secret Key***: Specify the Amazon Web Services Secret Key to use to access the data for download.
- In the drop-down, select an encryption option for the *AWS Secret Key*:

- ***Hide (Default)***: Hide the password using minimal encryption.
- ***Encrypt for Machine***: Any user on the computer will be able to fully use the connection.
- ***Encrypt for User***: The logged in user can use the connection on any computer.

Save Current AWS Credentials As Default: Saves the AWS credentials to the machine's registry.

Delete Saved Default AWS Credentials: Deletes any previously saved AWS credentials from the machine's registry.

- Endpoint: Select Default to allow Amazon to determine the endpoint automatically based on the bucket you select. To specify an endpoint for private S3 deployments, or if you know a specific bucket region, you can alternately select an endpoint (S3 region), enter a custom endpoint, or select from one of ten previously-entered custom endpoints.

Use Signature V4 for Authentication: Select this option to use Signature Version 4 instead of the default Signature Version 2. This will increase security, but connection speeds may be slower. This option is automatically enabled for regions requiring Signature Version 4.

- ***Bucket Name***: AWS stores data objects in Buckets. Type a Bucket name or select one from the list of available Buckets.
- ***Object Name***: Specify the Object name (data file) to be stored in the previously specified Bucket. Type an Object name or select one from the list of available Objects.

12.5 Amazon S3 Upload

<table>
<tr><td rowspan="2"> 
Figure 12.12 – Amazon S3 Upload</td><td colspan="3">The Amazon S3 Upload tool will transfer data from Alteryx to the cloud where it is hosted by Amazon Simple Storage Service.</td></tr>
<tr><td>Group</td><td>Input</td><td>Output</td></tr>
<tr><td></td><td>Connectors</td><td>Any data stream</td><td>None</td></tr>
<tr><td colspan="4">Write CSV, BDF and YXDB files from Amazon S3.</td></tr>
</table>

Properties Window:

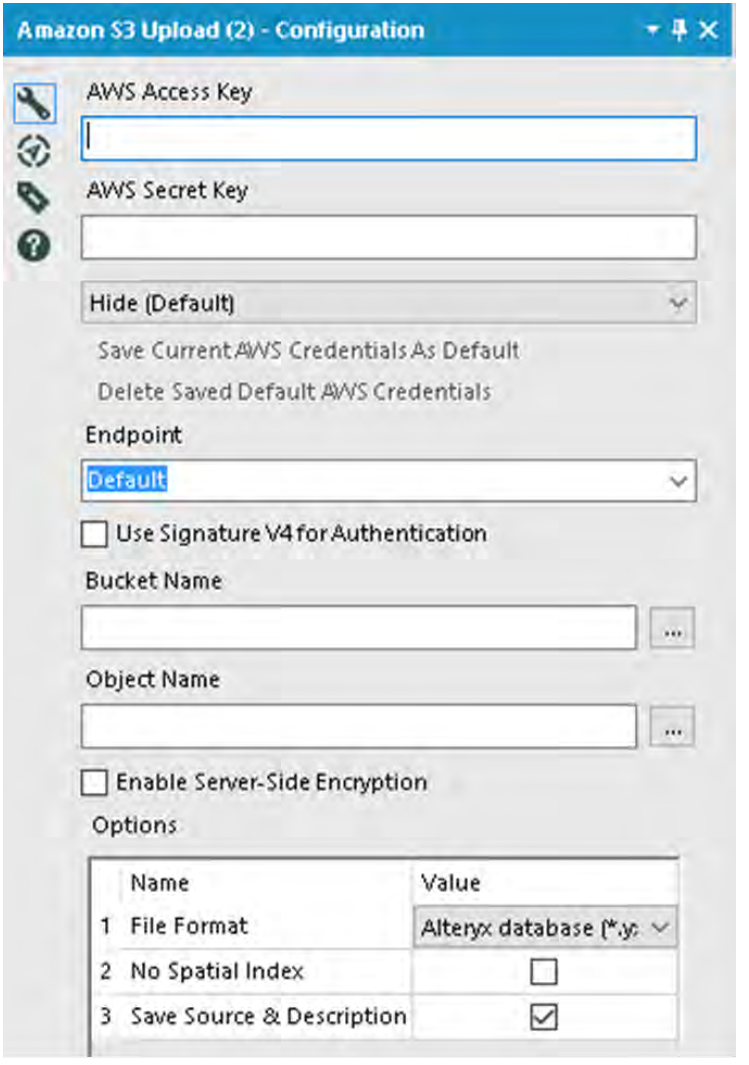


Figure 12.13– Amazon S3 Upload Configuration

- *AWS Access Key:* Specify the Amazon Web Services Access Key to use to upload data.
- *AWS Secret Key:* Specify the Amazon Web Services Secret Key to use to access the data for upload.

- In the drop-down, select an encryption option for the *AWS Secret Key*:
    - *Hide (Default)***:** Hide the password using minimal encryption.
    - *Encrypt for Machine***:** Any user on the computer will be able to fully use the connection.
    - *Encrypt for User***:** The logged in user can use the connection on any computer.

*Save Current AWS Credentials as Default***:** Saves the AWS credentials to the machine's registry.

*Delete Saved Default AWS Credentials***:** Deletes any previously *saved AWS credentials from the machine's registry.*

- *Endpoint:* Select Default to allow Amazon to determine the endpoint automatically based on the bucket you select. To specify an endpoint for private S3 deployments, or if you know a specific bucket region, you can alternately select an endpoint (S3 region), enter a custom endpoint, or select from one of ten previously-entered custom endpoints.

*Use Signature V4 for Authentication***:** Select this option to use Signature Version 4 instead of the default Signature Version2. This will increase security, but connection speeds may be slower. This option is automatically enabled for regions requiring Signature Version 4.

- *Bucket Name:* AWS stores data objects in Buckets. Type a Bucket name or select one from the list of available Buckets.

- *Object Name:* Specify the Object name (data file) to be stored in the previously specified Bucket. Type an Object name or select one from the list of available Objects.
- *Enable Server-Side Encryption:* Select this option to allow files to be uploaded to an encrypted Amazon S3 bucket. The only method supported at this time is SSE-S3.

12.6 Foursquare Search

Figure 12.14 – Foursquare Search

Search Foursquare Venues by location with an option to filter by a search term.

Group	Input	Output
Connectors	None	See Primary Output and Secondary Output

Note: Before you can use this macro, you must register an application with Foursquare. This can be set up by visiting https://foursquare.com/login and selecting 'Create a New App'. The only required fields are the 'Download/welcome page URL' and 'Redirect URL(s)', but http://www.foursquare.com is accepted as a placeholder in those fields https://foursquare.com/login. Once you have submitted your application form you will be provided a 'Client ID' and 'Client Secret' token, which are needed to configure the macro.

Primary output. This is the list of venues that fit the criteria selected in the interface, returned via the "P" macro output.

Secondary output. This optional output returns via the "S" macro output and contains either the NextVenues, Tips, or Photos if one of those options is selected in the interface; it will be empty if no secondary output is selected.

Properties Window:

There are 3 tabs: Authentication, Specification, Location

Authentication Tab

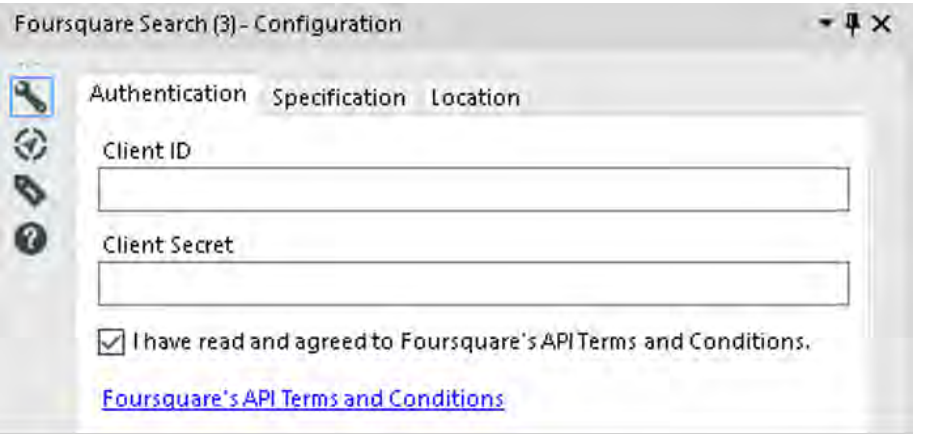


Figure 12.15– Foursquare Search Configuration

- *Client ID:* Enter your Foursquare Application Client ID.
- *Client Secret:* Enter your Foursquare Application Client Secret.
- *Foursquare API Terms and Conditions*: In order to run this macro, you must read and agree to the Foursquare API Terms and Conditions.

Specification Tab

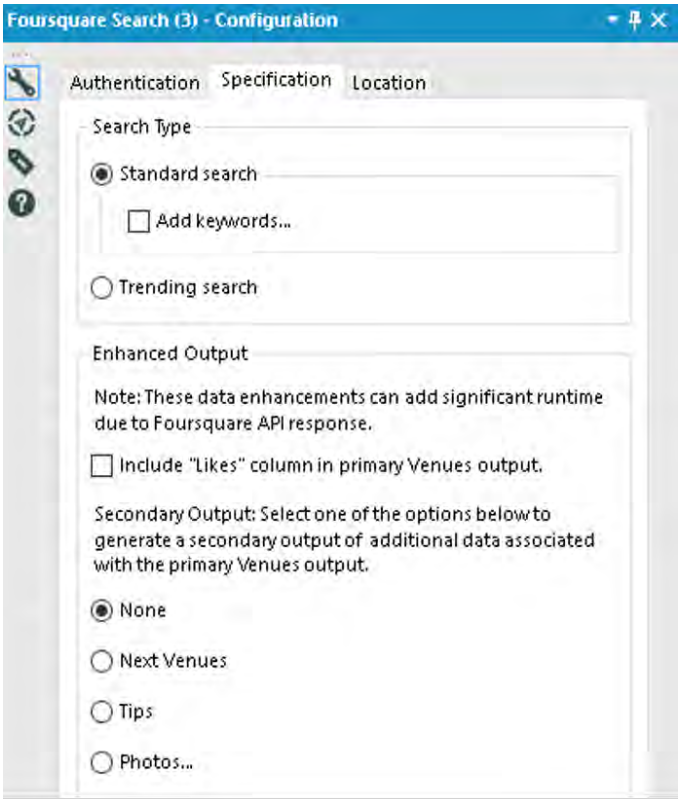


Figure 12.16– Foursquare Search Configuration

- Specify the *Search type*:
    - *Standard Search*: Returns a list of venues near the specified location.
    - *Add keywords*: When checked, the venues will be filtered to return only those relevant to the search terms entered.
    - *Keywords*: Specify keyword search terms, separated by commas. A single 'search term' can be 1 or more words, and a maximum of 10 search terms may be specified at a time. (e.g. burgers,ice cream,steak) Each comma-separated search term will have a maximum of

50 venues returned. (So in the previous example, you can get up to 150 records in total—50 for each search term.)
- *Trending Search*: Returns a list of up to 50 venues near the specified location with the most people currently checked in. (Keywords are not available in Trending searches.)

- Enhanced ***Output*:**
  - The Foursquare macro allows you to enhance the primary venue-related data returned, by enhancing the primary output and/or producing an additional secondary output. Be aware that selecting these options may—to varying degrees—add significant runtime to your process due to awaiting responses from the Foursquare API.
  - Include "Likes": Check this box to populate the Foursquare "Likes" column in the primary Venues output.

- *Secondary* ***Output*:** Select one of these options to generate a 'Secondary' output stream of data associated with the primary venues output. Join the secondary data output to the primary venues output by the FoursquareVenueID column that's found in both the primary and secondary output data streams. Choices include:
  - *None*: No secondary output is generated.
  - ***Next Venues*:** Returns venues that people often check into after the current (i.e. 'primary') venue. Up to 5 'next' venues are returned for each 'primary' venue.
  - ***Tips*:** Tips that members have posted to Foursquare.

- ***Photos*:** Photos related to the venue. Up to 200 photo records are returned for each venue.

Include actual image data: By default, the Photos output contains URL information and other metadata (username, photo width/height, etc) related to the photo, but not the actual image. Check this box to include the image. Including the actual image greatly increases both the time required to run the macro and the volume of data returned.

Location Tab

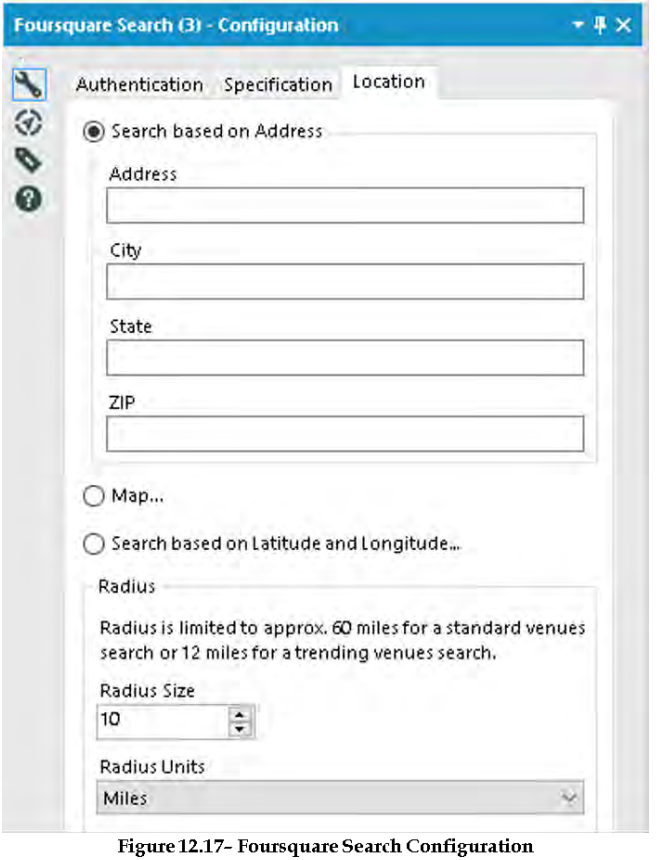


Figure 12.17– Foursquare Search Configuration

- **Enter *the search location* in 1 of 2 ways*:***

  - *Address*: Specify a Street Address, City, State and ZIP Code. To search based on an address, you must have licensed CASS and Geocoder.
  - *Latitude/Longitude*: Specify a latitude and longitude point.
  - *Radius Size*: Specify a radius for the venue search. Note: The maximum radius search supported by Foursquare is 100km (~60 miles) for a standard venues search or 20km (~12 miles) for a trending search. If the radius entered is larger, it will be automatically shortened to the appropriate limit.

- **Radius *Units*:** Specify Miles, Kilometers, Meters or Feet as the unit of measure of the radius.

12.7 Google Analytics

Figure 12.18 – Google Analytics

The *Google Analytics* tool downloads data from Google Analytics directly into your Alteryx workflow, allowing non-technical business users to utilize the Google Analytics API.

Group	Input	Output
Connectors	None	Any data stream

Output Data: The selected data from your query.

Output Summary: Summary information from your query.

Properties Window:

- **In** the *Configuration* **window**, select a sign-in method.

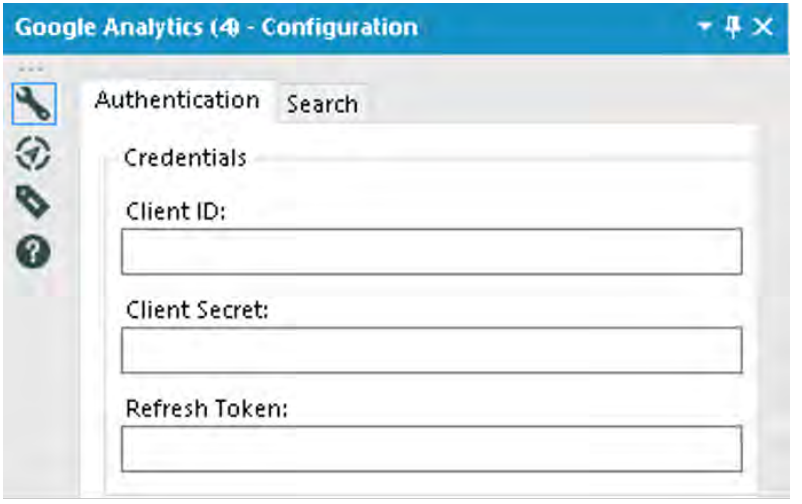


Figure 12.19– Google Analytics Configuration

  - *Online*: Use this method for ad-hoc workflows. Sign in using Google account credentials. This option requires re-entering credentials every 60 minutes or any time a new workflow is opened. This method will not work for scheduled workflows.
  - *Offline*: This method is required for scheduled workflows. Sign in using Google API credentials. This option requires a Client ID, Client Secret, and Refresh Token and does not require re-entering credentials to run a workflow.

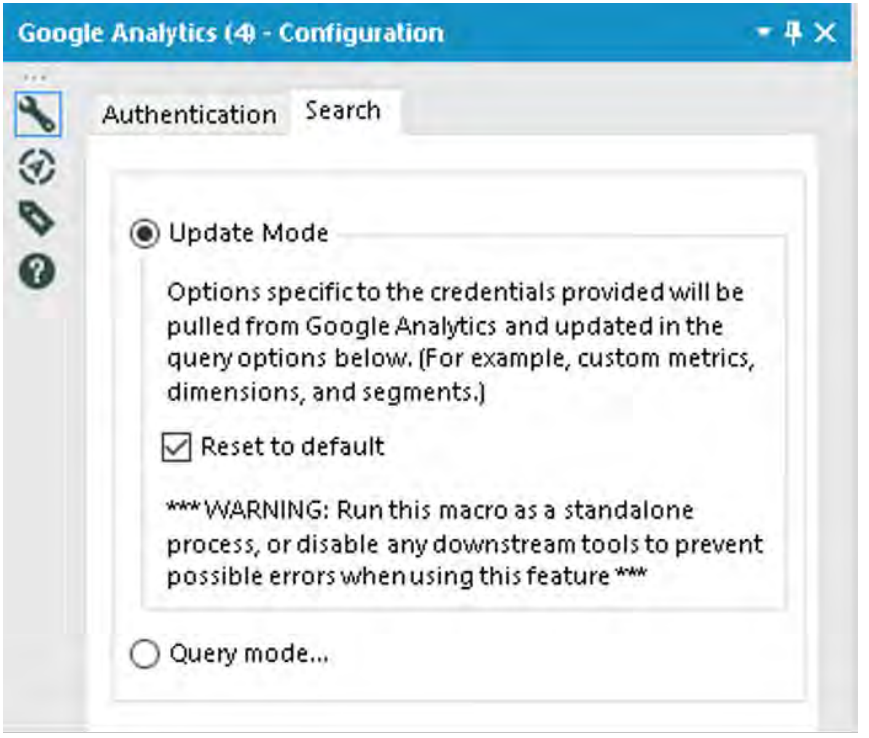


Figure 12.20– Google Analytics Configuration

- **Select** *an account*, *web property, and profile*.
    - *Available Accounts* is determined by your login account.
    - *Available WebProperties* is determined by your account selection.
    - *Available Profiles* is determined by your WebProperties selection. This is a unique table ID associated with the data query.

- **Select a *date range*.** Select either a preset range (for example, Today, Last Month, Year to Date) or Custom to specify a start and end date.

- **Select at least one *metric* and a maximum of ten metrics and goals. combinations (Optional) Select a maximum of seven dimensions**.

Dimensions break down metrics by common criteria. Only valid combinations of metrics and dimensions can be used.

- **(Optional) Select a maximum of four segments.** Selecting multiple segments limits the results to data included in all selected segments.

12.8 Marketo Append

Figure 12.21 – Marketo Append

The *Marketo Append* tool retrieves Marketo records and appends them to the records of an incoming data stream.

Group	Input	Output
Connectors	Any data stream	Any data stream

Using the Append tool would be most helpful in situations where you have a list of email addresses or other data that you are already working with, and want to join that data with additional information from Marketo, or see if they already exist in your Marketo instance

Note: All dates in Marketo are stored with a UTC offset. This format is comprised of the local time with an appended offset that can be positive or negative.

Properties Window:

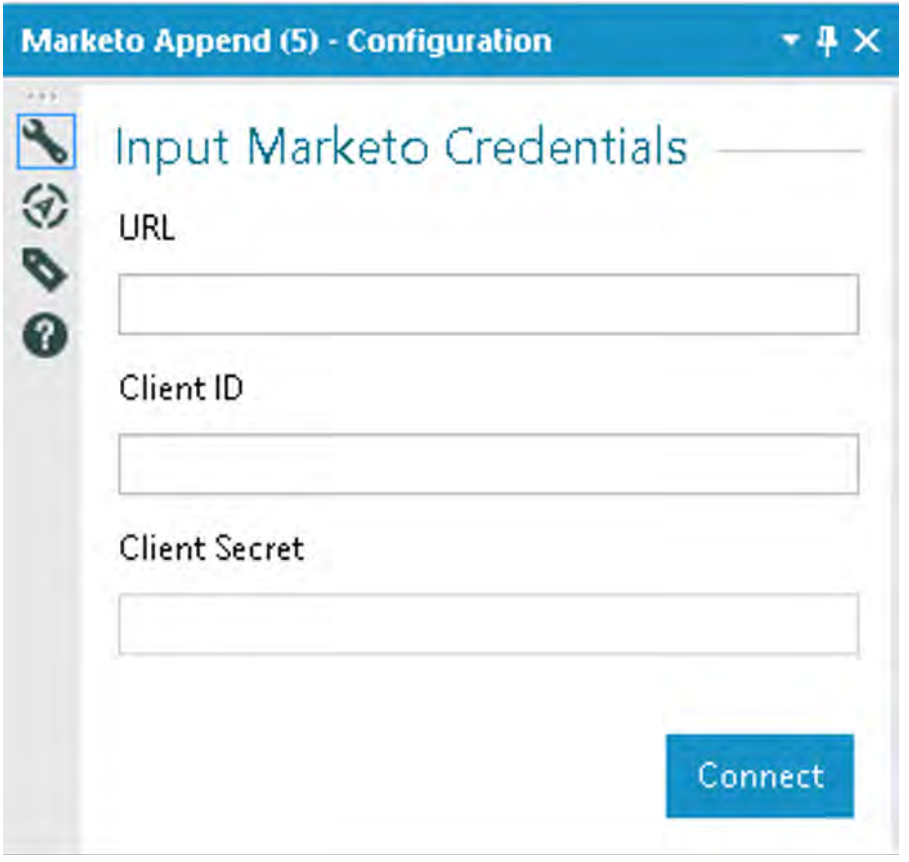


Figure 12.22– Marketo Append Configuration

- *URL:* The Marketo REST Instance to retrieve records from. This information is found in Marketo (Admin > Integration > Web Services > REST API).
- *Client ID:* Client ID for the API Role user. This information can be found in Marketo (Admin > Integration > LaunchPoint).

- *Client Secret:* Client Secret for the API Role user. This information can be found in Marketo (Admin > Integration > LaunchPoint).
- *Incoming Field:* This drop-down will show the list of fields that you have used as input for the tool. Marketo will use this field to find lead records with this field value in the Marketo instance.
- *Incoming Field Type*: What type of records the incoming field is (email, SFDC id, etc.).
- *Output Fields*: The fields you want to retrieve from the Marketo database. These fields, as well as the fields that were passed into the tool, will be returned in the output data stream. If the output fields that you select are also contained in the incoming data stream, the output field coming from Marketo will be renamed with a '2' at the end.

12.9 Marketo Input

<table>
<tr><td rowspan="2"> 
Figure 12.23 – Marketo Input</td><td colspan="3">The Marketo Input tool reads Marketo records for a specified date range.</td></tr>
<tr><td>Group</td><td>Input</td><td>Output</td></tr>
<tr><td></td><td>Connectors</td><td>None</td><td>Any data stream</td></tr>
<tr><td colspan="4">Use the Marketo Input tool if you do not have any data to start with, or to sample data from your Marketo instance. If you have data that you want to join to additional data from Marketo, the Marketo Append tool is a better option</td></tr>
</table>

Two types of Marketo records can be retrieved:

LeadRecords: These are lead records and there will be one record for each lead.

Lead Activity Records: These records track the activities for each lead. There are potentially many Lead Activity Records for each Lead Record.

The Properties Window is the same as for Marketo Append: refer to Section 12.8

Based on the selections of the type of record, different options are available and based on that selection there will be parameters available to configure, some are optional and some are required. The table below details which parameters are required based on your chosen configuration.

Choose parameters

- *Lists*: All static lists for the Marketo instance specified in the configuration are listed here (smart lists are not currently available via the API)-only 1 list selection is allowed per request.
- *Activity Types*: All activity types available in Marketo are listed here, there is a maximum of 10 activity type selections per request.
- *Output Fields*: All available fields available in the Marketo instance will be listed. If no selection is made the default list of fields will be returned: Email, Lead ID, First Name, Last Name, Updated At, Created At.
- *Start Date/Time*: Specifies the starting point in the database to begin retrieving records.

Note: Error messaging such as authentication failures, etc. will be displayed in the configuration window. Error messaging such as missing configuration selections will be shown over the tool in the workflow.

The tool sends records in batches of 300 records or less, each batch counts as a separate API call towards the daily limit set by Marketo.

The Marketo REST API limits accounts to 10,000 requests per day, requests to the REST API do not count towards the SOAP API daily limitations.

You want to make sure to maintain Marketo field structures in Alteryx, before writing back to Marketo to avoid errors. The Marketo connector will convert the data to the proper datatypes before outputting to Marketo as long as the field names are valid. If the field names are not found in Marketo, the output will fail.

12.10 Marketo Output

Figure 12.24 – Marketo Output

The *Marketo Output* tool helps to write back data to Marketo using an 'Upset' operation.

Group	**Input**	**Output**
Connectors	Any data stream	None

The Marketo Output tool makes a call to the Marketo REST API endpoint: Create/Update Leads. Data is output to Marketo based on the action you select in the configuration.

The Properties Window is the same as for Marketo Append: refer to Section 12.8

- *Key Field*: This drop-down will show the list of fields that you have used as input for the tool. Marketo will use this field to find any duplicate records. If no field is specified, then the email will be used by default. If the field 'id' is included in your records, 'id' must be selected as the key field.
- *Output Action*
  - Create or Update (default): Based on the key field, create a lead if the lead does not exist, update the lead if it does exist.
  - *Create Duplicate*: Create another lead even if it already exists.
  - Update Only: Update the lead if it already exists, do nothing if it does not.
  - *Create Only*: Create a new lead if it does not exist, if it does exist, skip it.
- *Partition Name*: If the Marketo Instance being accessed has partitions set up, they will be listed in the drop-down. This field selection is required if the instance being written out to has lead partitions set up. If there are no partitions available, 'Default' will be used.

12.11 MongoDB Input

Figure 12.25 – MongoDB Input

The *MongoDB Input* tool reads data stored in MongoDB databases.

Group	Input	Output
Connectors	None	Any data stream

The MongoDB Input tool is used for reading data stored in MongoDB databases. MongoDB is a scalable, high-performance, open source, NoSQL database. You can learn more about MongoDB here: http://www.mongodb.org/.

MongoDB databases store data in a BSON format which is binary JSON. You can learn more about BSON here: http://bsonspec.org/.

Properties Window:

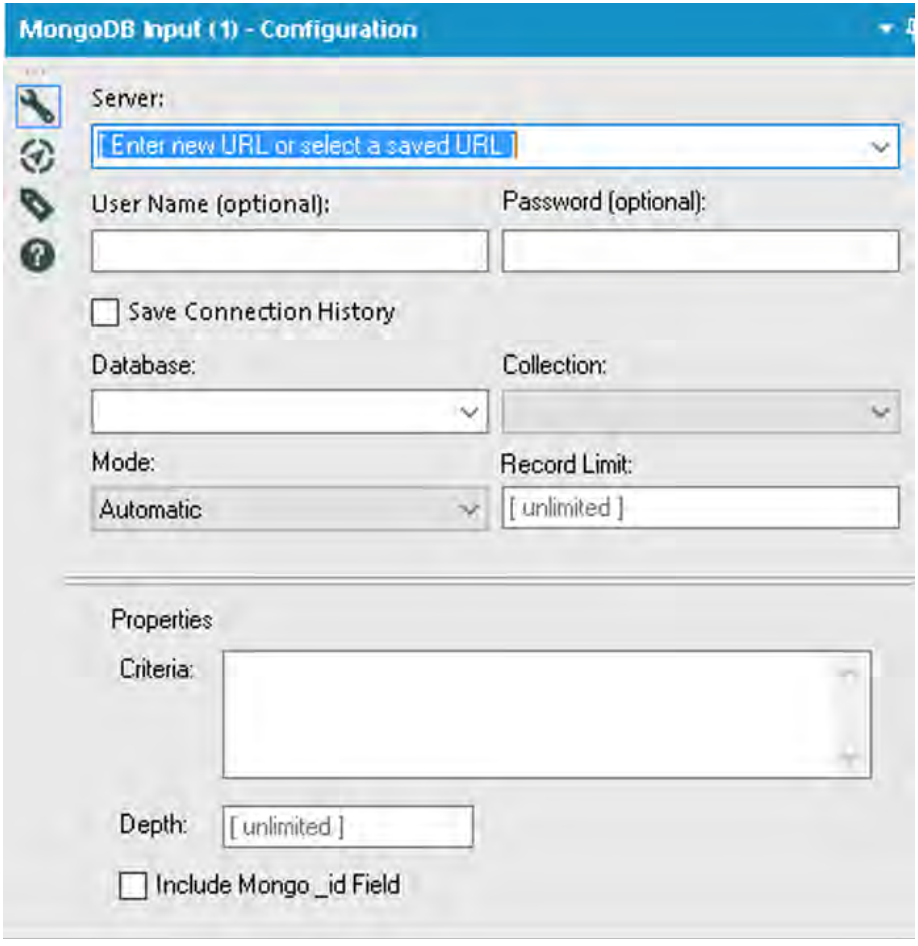


Figure 12.26– MongoDB Input Configuration

- *Server*: The name of the MongoDB server you wish to connect to. Enter localhost to connect to a MongoDB instance you have on the machine running Alteryx.
- *User Name/Password* (optional): If your MongoDB instance is running with the –auth option this is where you enter the username/password you wish to connect to.

- Database: The name of the MongoDB database you wish to connect to.
- *Collection*: The name of the MongoDB collection you wish to read data from.
- *Mode*: There are 2 modes for reading the MongoDB. Depending on the mode chosen, additional configuration is necessary. Choices include:
    - *Automatic*: Alteryx will read the data in 2 passes: the first pass will scan through all documents to determine the table schema (columns, data type, and size); the second pass will return the data.
    - *Manual*: In Manual mode, you specify the schema or you can choose to scan a specified amount of documents to determine the schema.

12.12 MongoDB Output

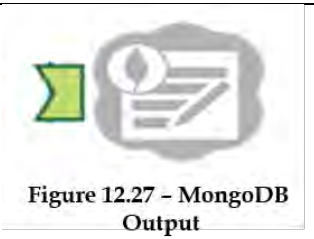


Figure 12.27 – MongoDB Output

The *MongoDB Output* tool writes data to MongoDB databases.

Group	Input	Output
Connectors	Any data stream	None

MongoDB is a scalable, high-performance, open source NoSQL database. You can learn more about MongoDB here: http://www.mongodb.org/.

MongoDB databases store data in a BSON format which is binary JSON. You can learn more about BSON here: http://bsonspec.org/.

Properties Window:

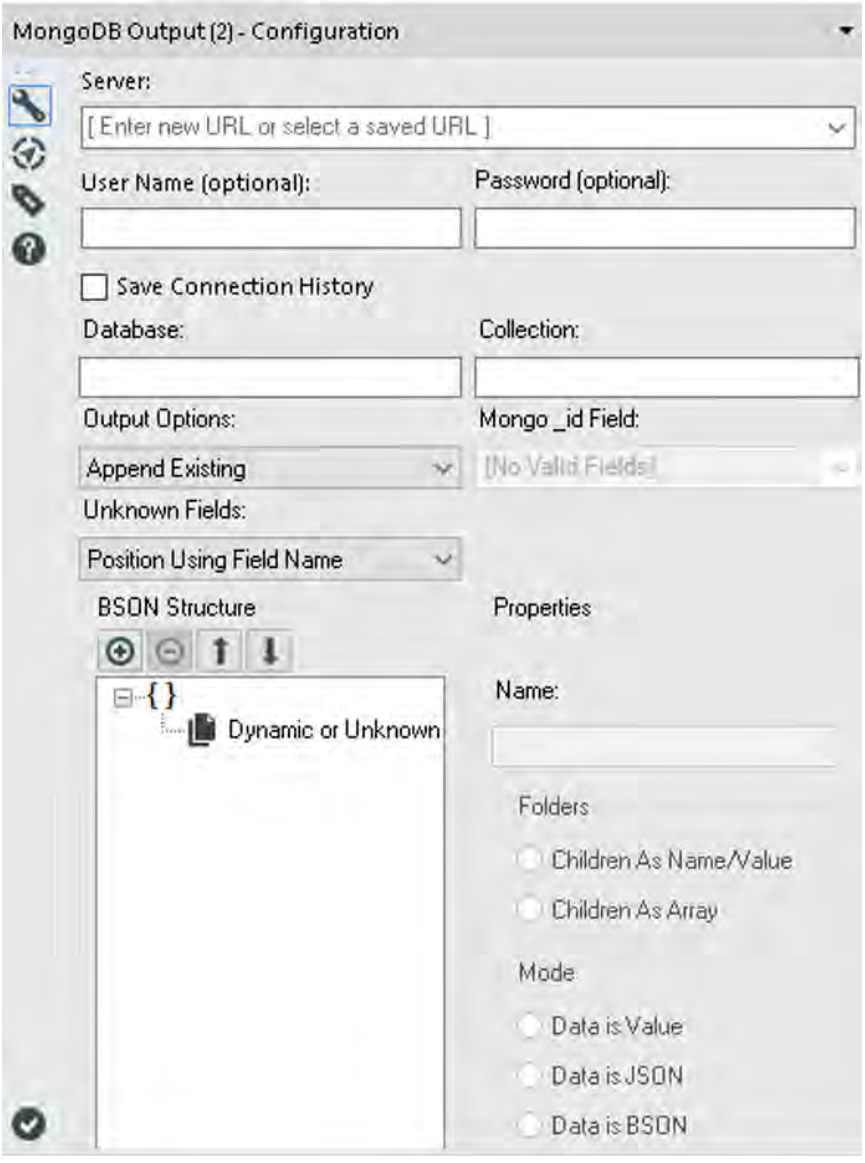


Figure 12.28- MongoDB Output Configuration

- *Server:* The name of the MongoDB server you wish to connect to. Enter localhost to connect to a MongoDB instance you have on the machine running Alteryx.
- *Database:* The name of the MongoDB database you wish to connect to.
- *Collection*: The name of the MongoDB collection you wish to push data to.
- *User Name/Password (optional)*: If your MongoDB instance is running with the –auth option this is where you enter the username/password you wish to connect to.
- *Output Options*:
  - *Append Existing*: Append the new records onto the end of your collection
  - *Delete Data & Append*: Deletes the collection and then adds the new records
  - *Update Using _id*: Will attempt to match an existing record using the MongoDB _id element to the field specified in "_id field". If a match is found then the existing record will be updated with the new one. If no match is found then the new record will be appended onto the collection.
- *Mongo_id Field*: Used to select the Alteryx field which contains the _id value when using "Update Using _id" mode. Note if you wish to update based on the MongoDB generated ID then this should be a JSON object e.g. { "_id" : { "$oid" "4fad55603346998a9f7d6841" }}
- *BSON view*: The tree view at the bottom of the tool configuration provides a visual representation of what the BSON object will look like which is going to be inserted into your collection.

  The default options use the dot notation to build the structure of the BSON object so, for example, passing fields Member ID; Name. First; Name. Last would by default give you a JSON object which looked like this {

Member ID: 123, Name: { First: "Alistair", Last : "Terry" } }. However, you can restructure the BSON object to look however you would like by using the up/down/add/remove buttons.

- *Name :* Allows you to rename any given BSON element (Note: BSON names cannot begin with $ nor contain '.'s)
- *Folders:*
  - *As Name/Value*: Will create child nodes as name value elements e.g. Name : { First : "Alistair" , Last: "Terry" }
  - *As Array*: Will add child nodes as member of an array e.g. Name: ["Alistair" , "Terry"]
- *Mode :(*for the incoming Alteryx fields)
  - *Data is Value*: The data in the Alteryx field is plain data to add to the data element of the BSON pair.
  - *Data is JSON*: The data in an Alteryx string field is already a JSON and should be appended as such.
  - *Data is BSON*: The data in an Alteryx blob field is already a BSON and should be appended as such.
- *Unknown Fields:*
  - *Position Using Field Name*: The unknown fields will be positioned using the dot notation of their names. Starting with the location of the Dynamic or Unknown Fields node as their root position.
  - *Position as Left*: The unknown fields will be added wherever the Dynamic or Unknown Fields node is positioned.

12.13 Salesforce Input

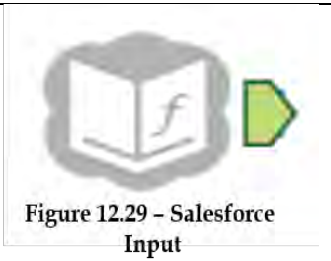


Figure 12.29 – Salesforce Input

The *Salesforce Input* tool allows you to read and query tables from Salesforce.com into Alteryx.

Group	Input	Output
Connectors	None	Any data stream

This tool requires that your Salesforce account is "API Enabled". Contact your Salesforce administrator for assistance with granting your account API user permissions.

Use Query Builder to browse for and select a Salesforce table, output fields, and other parameters. Use Custom Query to specify a table and parameters as a Salesforce Object Query Language (SOQL) query.

Properties Window:

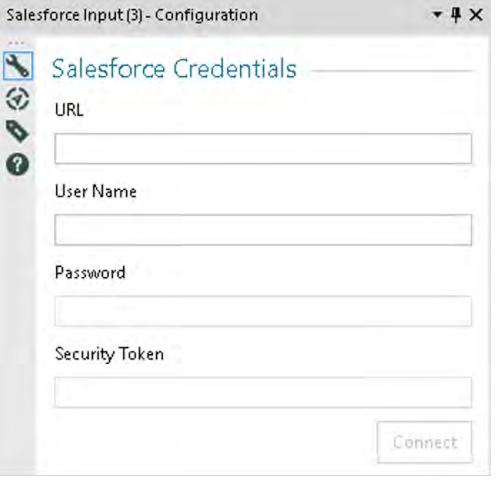


Figure 12.30– Salesforce Input Configuration

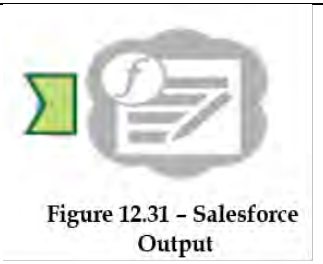
- *URL*: Enter your Salesforce URL in the following format: https://[instance].salesforce.com where [instance] refers to the specific server that pertains to your Salesforce environment. (For example https://na9.salesforce.com). The easiest way to determine your instance is to log into Salesforce in a browser at https://login.salesforce.com. Once you are logged in you will be redirected to a URL that contains your instance.
- *User Name*: Enter your Salesforce username associated with the Salesforce URL specified above. This is often an email address.
- *Password*: Enter your password for the Salesforce username specified above. This information will be encrypted.
- *Security Token*: You may need to enter your Salesforce Security Token. This information will be encrypted. You can find instructions on how to get your security token from Salesforce Help.
- *Connect*: Click this button to establish a connection. If the credentials are accurate, the Configuration window will display.
- *Query Builder*
  - *Table*: Click a table to read into Alteryx. Type in the search box to filter for a table or browse for a table by using the arrows – but you must click on a table name to register your selection. This list will only include queryable tables. Queryable here is a reference to a flag returned from the API. If this is set to false, then those tables will not be displayed in the list.
  - *Output Fields*: Select the fields you wish to retrieve from the above table. When no fields are selected, all output fields will be returned. This field is optional.

- *Record Limit*: Enter a number of rows to return. If left blank, all rows will be returned. This field is optional.
- *WHERE Clause (SOQL):* Specify a Query on the table specified above. Query language must be SOQL (Salesforce Object Query Language). This field is optional.

- SOQL Query: Specify a custom SOQL statement in the text box. If you previously selected a table using the Query Builder option, you will be prompted to load that query into Query Builder as a starting point.
  - Attempt to Parse JSON Response: When selected, Alteryx will attempt to parse the query response and will display the parsed output in the Results window. If not selected, the response is output as a single field (called 'JSON'), which you can then parse with the JSON Parse tool.
  - Validate: Click this button to submit the query to the Salesforce API to determine if the query is valid. Alteryx will also determine if the query results can be parsed. If Alteryx cannot parse the results, the response will be output as a single field (called 'JSON'), which you can then parse with the JSON Parse tool.

12.14 SalesForce Output

	The *Salesforce Output* tool allows you to write to Salesforce.com tables from Alteryx.		
	Group	**Input**	**Output**

 Figure 12.31 – Salesforce Output	Connectors	Any data stream	None
This tool requires that your Salesforce account is "API Enabled". Contact your Salesforce administrator for assistance with granting your account API user permissions Note: Do not use Blob and SpatialObj field types in your workflow, as they cannot be output to Salesforce.			

Properties Window:

The SalesForce Output has the same configuration window as SalesForce Input, where you need to fill Salesforce credentials.

- *Table*: Click a table to write to from Alteryx. Type in the search box to filter for a table or browse for a table by using the arrows – but you must click on a table name to register your selection.
- *Output Operation*: When writing data to a Salesforce.com table, the field names must be the same as the field names in the table, including capitalization, but the type and size of the data do not matter. Choose from the following output options:
- *Update*: Replaces existing records in the specified table with the contents of the input stream. The ID field is required. All other fields must be within the target table.

- *Insert*: Adds to existing records in the specified table with the contents of the input stream. The ID field cannot be included. All other fields must be within the target table.
- *Delete*: Removes the records in the input stream from the specified table. The ID field is required and should be the only field provided.
- *Change Credentials*: Click the link at the bottom to change connection details, if necessary.

12.15 SharePoint List Input

Figure 12.32 – SharePoint List Input

The *SharePoint Input* tool reads lists from SharePoint to be used as a data input in a workflow.

Group	Input	Output
Connectors	None	Any data stream

SharePoint automatically stores date/time data in UTC format. Date/time data is automatically converted to your local timezone when Alteryx reads a SharePoint list.

Properties Window:

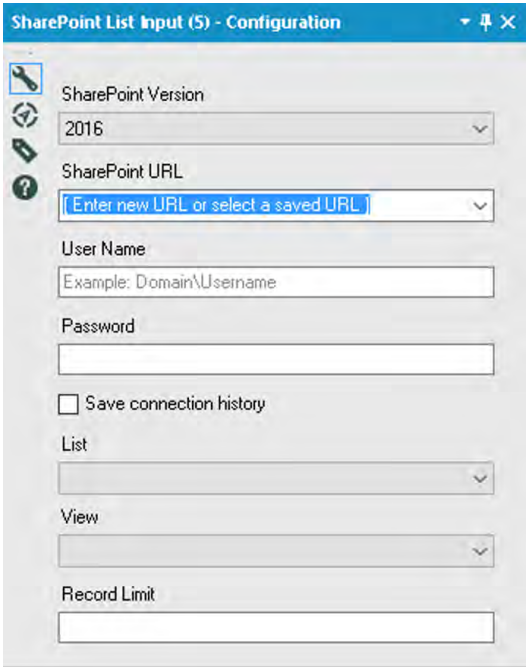


Figure 12.33– SharePoint List Input Configuration

- Select a *SharePoint Version:* 2007, 2010, 2013, 2016, or Online.
- Type the full *SharePoint URL* or click the drop-down to select a saved URL.
- Type your *User Name and Password*. The user name must include the domain name.
- (Optional) Select *Save connection history* to save the connection URL and credentials when you run the workflow. You can save up to ten connections.
- Select a *List* from the drop-down. The List drop-down contents are specific to the URL you specify. If your URL contains a directory, only lists for that directory are shown.
- Select a *View* from the drop-down. The available views are determined by the list you select.

- (Optional) Specify a *Record Limit*. To read in all records, leave this field blank.

12.16 SharePoint List Output

Figure 12.34 – SharePoint List Output

The *SharePoint output* tool writes the content of a data stream to a Sharepoint list.

Group	Input	Output
Connectors	Any data stream	None

Note: SharePoint automatically stores date/time data in UTC format. Date/time data is automatically converted to UTC when Alteryx writes a SharePoint list

Properties Window:

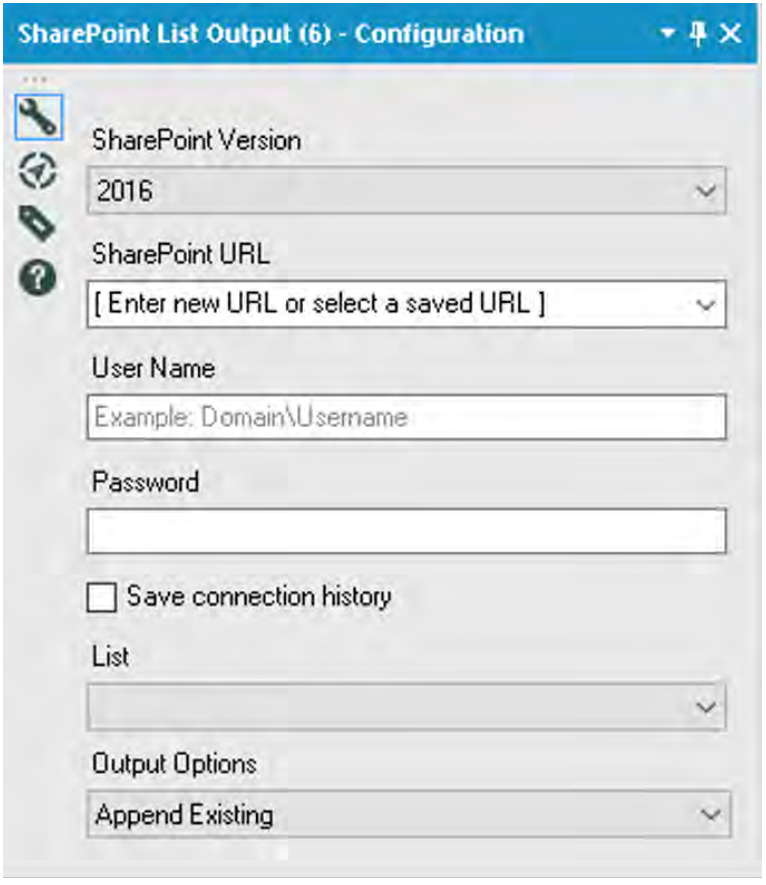
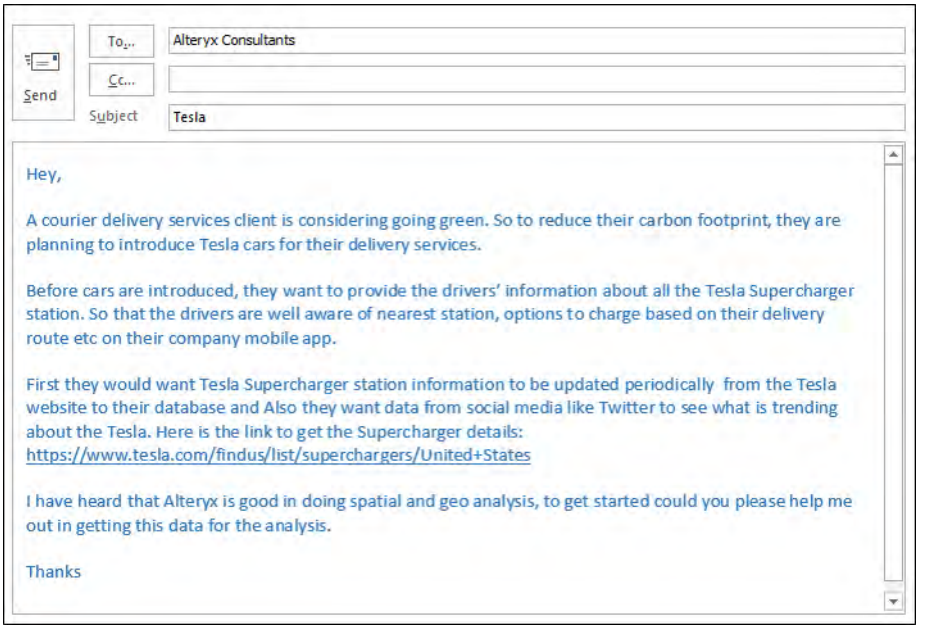


Figure 12.35– SharePoint List Output Configuration

- *Select a SharePoint Version*: 2007, 2010, 2013, 2016, or Online.
- Type the full *SharePoint URL* or click the drop-down to select a saved URL.
- Type your *User Name and Password*. The user name must include the domain name.
- (Optional) Select *Save connection history* to save the connection URL and credentials when you run the workflow. You can save up to ten connections.

- Select *a List* from the drop-down. The List drop-down contents are specific to the URL you specify. If your URL contains a directory, only lists for that directory are shown.
- In *Output Options*, select an option for writing the list:
  - *Append Existing*: Appends all the data to an existing table.
  - *Delete List and Append*: Deletes all the original records from the table and then appends the data into the existing table.
  - *Update; Warn on Update Failure*: Updates only existing records with the IDs you specify. If a record cannot be updated, a warning is reported.
  - *Update; Error on Update Failure*: Updates only existing records with the IDs you specify. If a record cannot be updated, an error is reported and processing will stop.

12.17 Green on the go



Based on the request to create the supercharger station data for the spatial and geo analysis, we will start looking at how to pull the information which is available on the Tesla website.

First, let us list down what we should be doing to extract the supercharger station data for the website in a specific format.

1. Get the URL of Tesla supercharger station information web page.
2. Look at the webpage, see what data would you extract for analysis. Like,

   Supercharger Station Name, Street Address, State, Zip and Roadside assistance

3. Download the page on Alteryx.
4. Parse the data from the above data points
5. Prepare and filter the data

We already know the URL where we can find the Tesla supercharger station information

URL:https://www.tesla.com/findus/list/superchargers/United+States

Let us use the URL in Text Input and pass the URL to the Download tool.

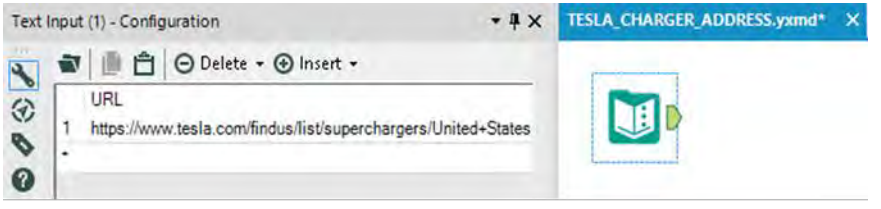


Figure 12.36- Green on the go - Text Input

Download tool takes the URL as input and retrieves data from a specified URL. We will configure the download tool to output data as a string. This option returns the data as a new wide string type field. A wide string supports Unicode characters.

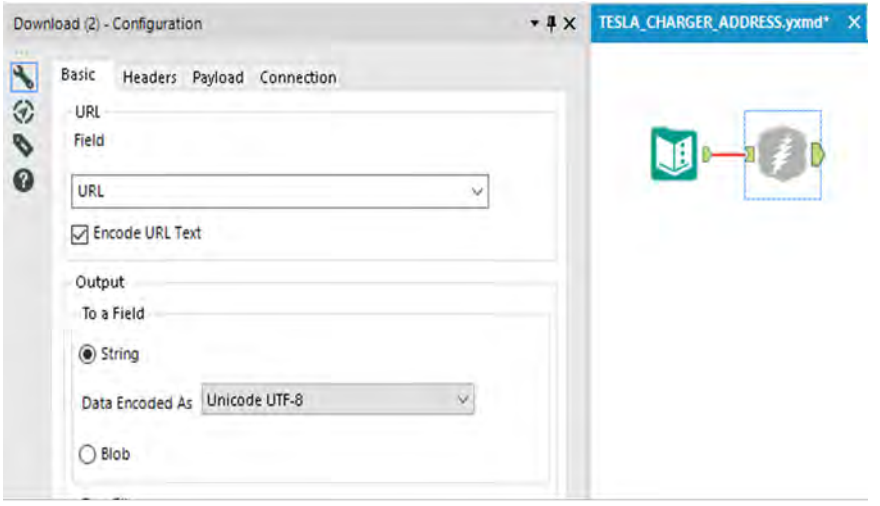


Figure 12.37– Green on the go – Download configuration

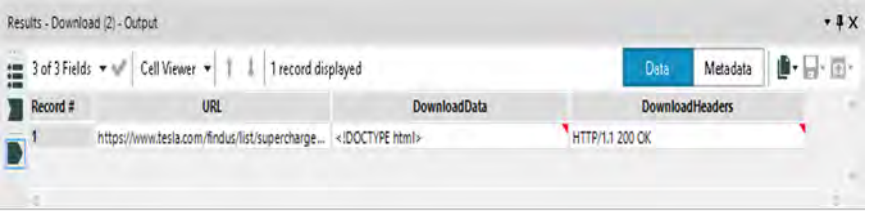


Figure 12.38– Green on the go – Download output

So now we have the result in a string or the text. Next step is to convert the string or the text to rows. We use the Text to column or row tool for this task.

At this stage running process shows us the following,

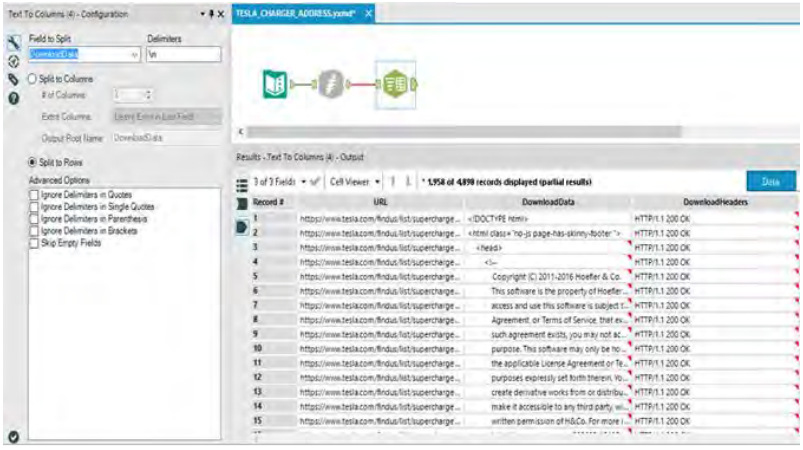


Figure 12.39– Green on the go – Text to column configuration and output

After converting the data to rows, the challenge is to get rid of the unwanted rows which were part of the string/text converted from the URL.

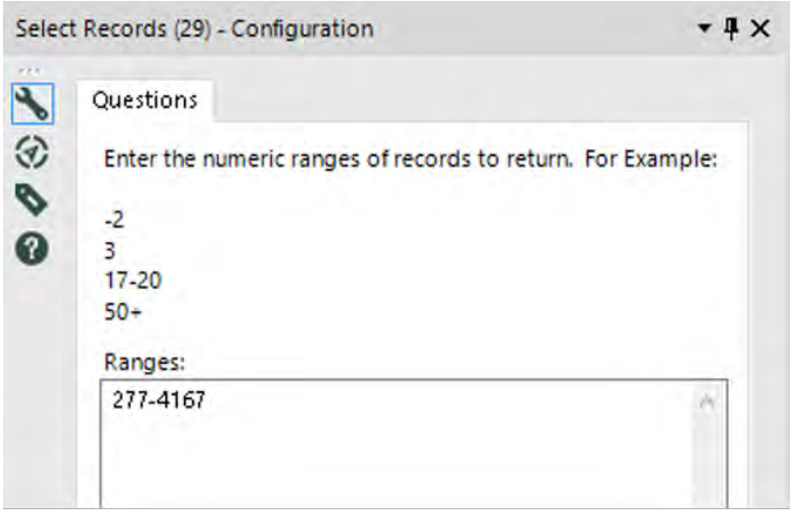


Figure 12.40– Green on the go – Select Records configuration

Using the select record, we will select the records from the row number 277-4167. Looking at the previous output we found the

first 276 rows does not contain the information that we want. So, we are filtering them out for the data stream.

Now is time to clean the data further. Using the data cleansing tool, let us do few cleansing operation.

Like,

- Replacing NULL with Blank for the string data
- Replacing NULL with 0 for the numeric data
- Removing unwanted characters like leading and trailing whitespaces, Tabs, Line Breaks etc.

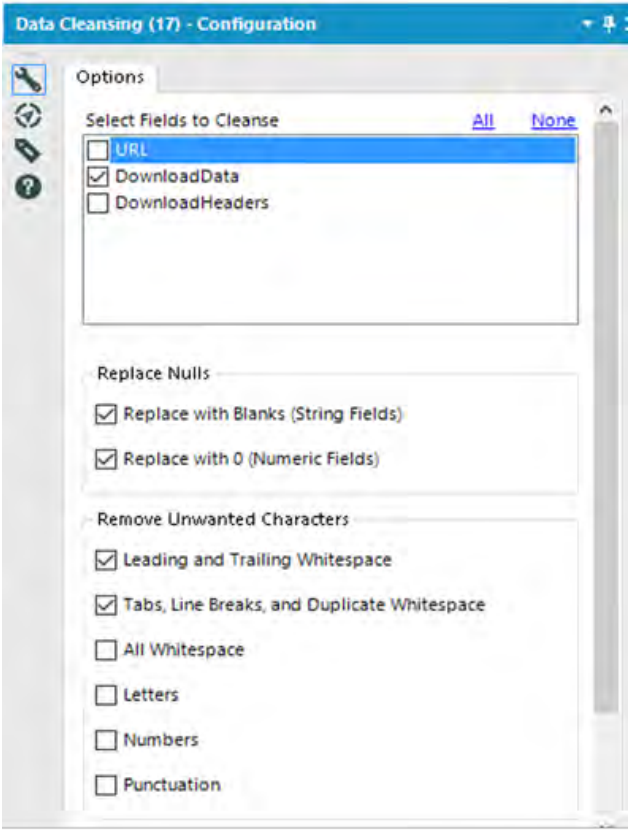


Figure 12.41- Green on the go - Data Cleansing configuration

Now we got a lot cleaner data.

Next step would be to create a mapping Create a mapping table using text input. Mapping table will have the data nodes that we would like to extract.

Roadside Assistance:

The above data nodes are mapped to the data points,

Supercharger Station Name
Street Address
State & Zip
Roadside assistance

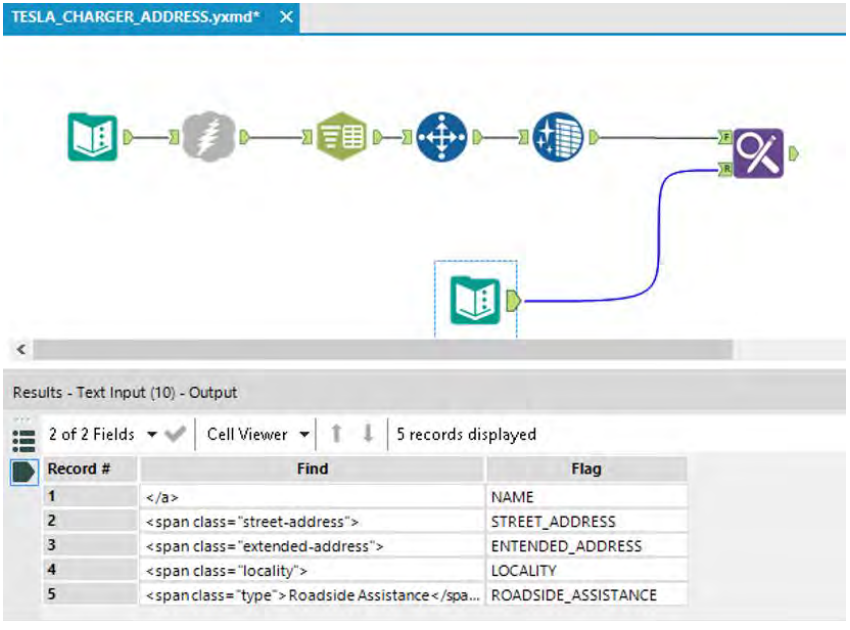


Figure 12.42– Green on the go - Text input

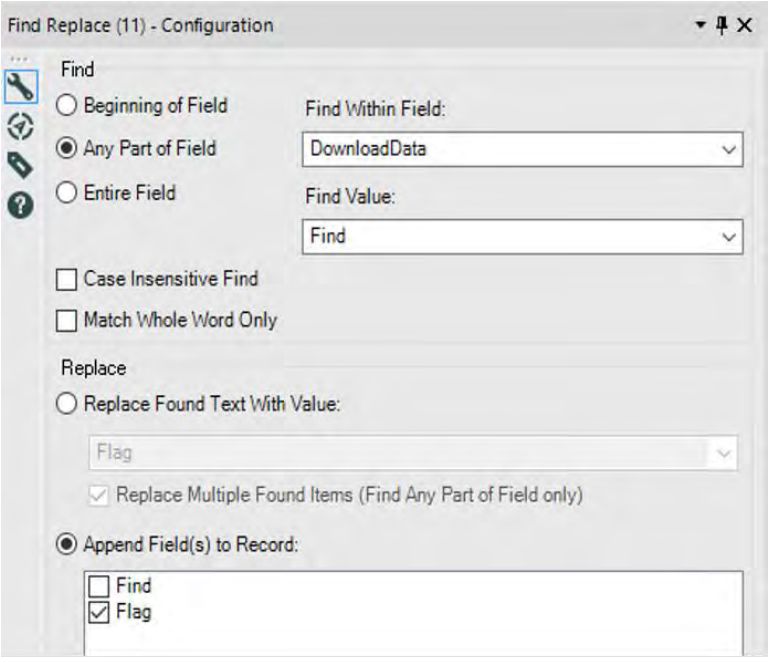


Figure 12.43– Green on the go –Find Replace Configuration

Filter the rows containing the data for the data point mentioned above.

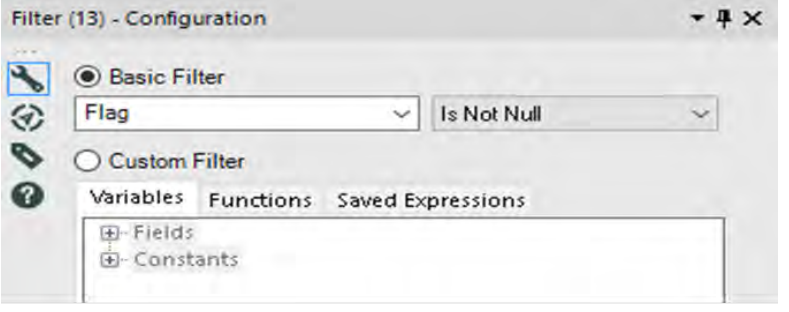


Figure 12.44– Green on the go –Filter Configuration

We have the data for Supercharger Station Name, Street Address, State & Zip and Roadside assistance

Since we have all the information which was requested, now time to clean up the unwanted HTML tags in the row and assign the unique ID to each Supercharger Station records.

Next 2 tools Multi-Row Formula and Formula tools will help us in assigning a unique number with a new column ID and removing unwanted HTML tags.

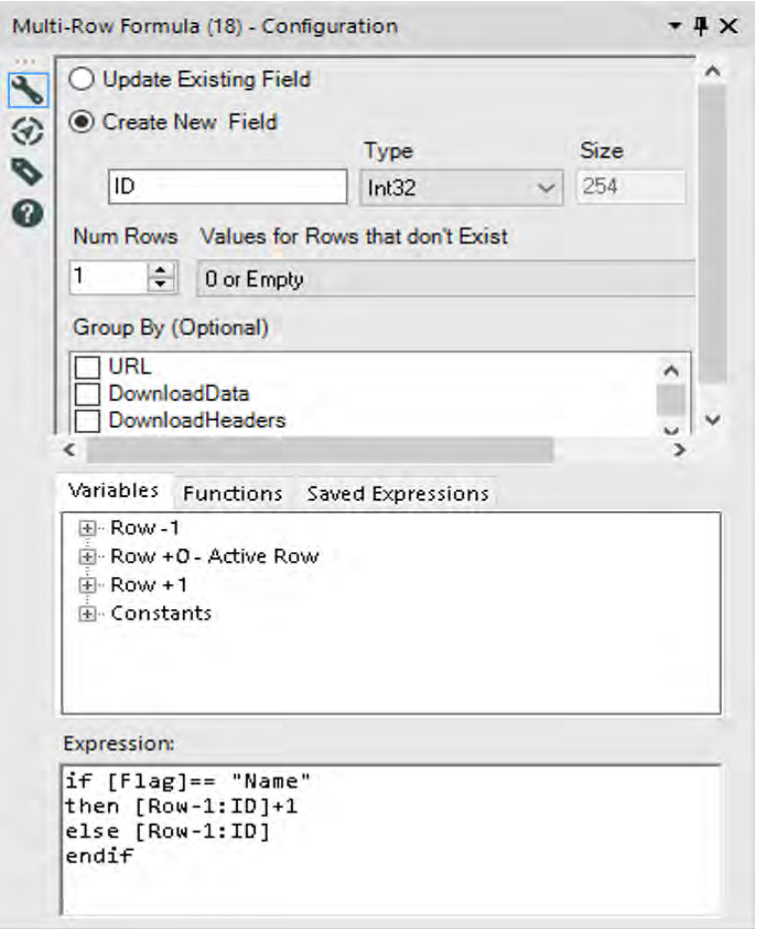


Figure 12.45– Green on the go –Multi-Row Configuration

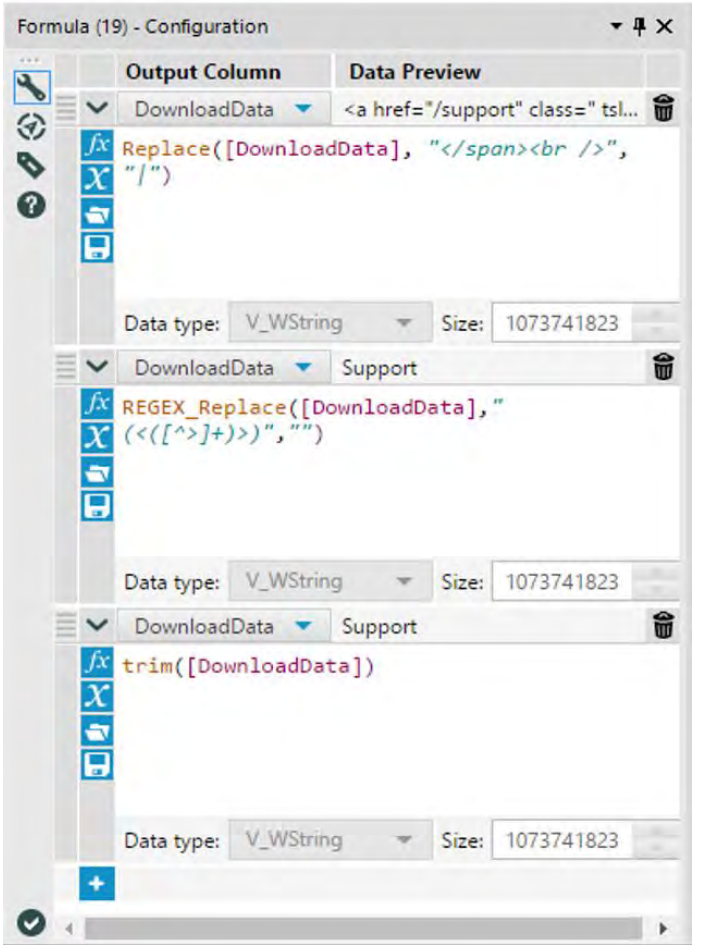


Figure 12.46– Green on the go –Formula Configuration

Till now we were massaging Supercharger Station data which were in rows. So now we would like to convert the relevant information as a column.

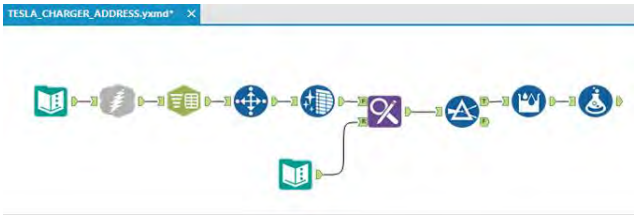


Figure 12.47– Green on the go –Data Stream

ID column which we just generated will help us to do a group for Supercharger Station and convert other information like the address, state, and zip into respective columns. We will use Cross Tab tool to reshape the data stream.

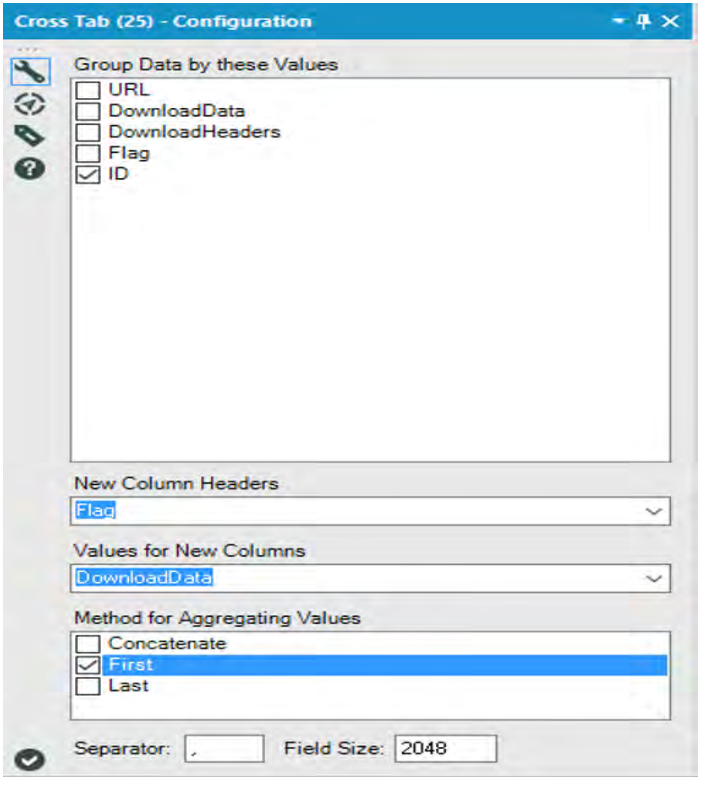


Figure 12.48– Green on the go –Cross Tab Configuration

The last step is to get rid of the unnecessary column and renaming the output column as required.

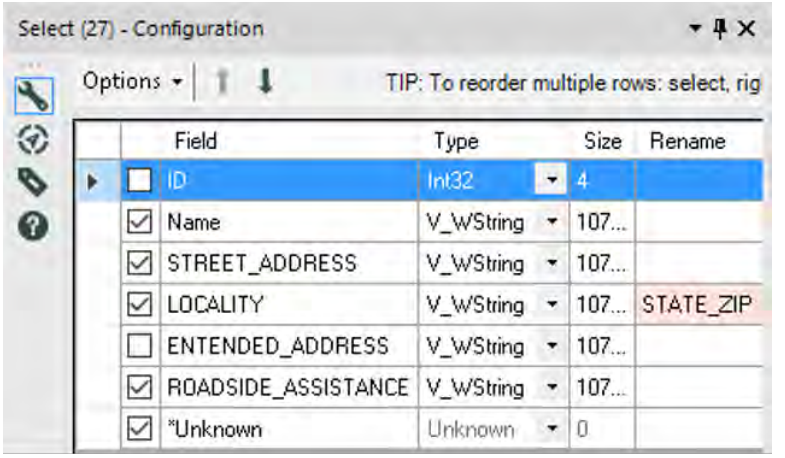


Figure 12.49– Green on the go –Select Configuration

One last time, make sure that there are no NULL rows in the data stream. We find first 7 rows are showing NULL values so we would like to filter out them for the data stream.

317 records displayed, 4 fields, , 20 KB

Table | Profile

4 of 4 Fields | Cell Viewer

Record #	NAME	STREET_ADDRESS	STATE_ZIP	ROADSIDE_ASSISTANCE
1	Support	[Null]	[Null]	[Null]
2	Enterprise	[Null]	[Null]	[Null]
3	Find Us	[Null]	[Null]	[Null]
4	Events	[Null]	[Null]	[Null]
5	Shop	[Null]	[Null]	[Null]
6	My Tesla	[Null]	[Null]	[Null]
7	Back to list	[Null]	[Null]	[Null]
8	Auburn Alabama Supercharger	1627 Opelika Road	Auburn, AL 36830	Roadside Assistance: (877) 798-3752|
9	Birmingham, AL Supercharger	2221 Richard Arrington Junior Blvd	Birmingham, AL 35203-1103	Roadside Assistance: (877) 798-3752|
10	Greenville Supercharger	219 Interstate Drive	Greenville, AL 36037	Roadside Assistance: (877) 798-3752|
11	Mobile Supercharger	3201 Airport Blvd	Mobile, AL 36606	Roadside Assistance: (877) 798-3752|
12	Buckeye Supercharger	416 S Watson Rd	Buckeye, AZ 85326	Roadside Assistance: (877) 798-3752|
13	Casa Grande Supercharger	2453 E. Florence Blvd.	Casa Grande, AZ 85194	Roadside Assistance: (877) 798-3752|
14	Cordes Lakes Supercharger	14925 Cordes Lakes Rd	Mayer, AZ 86333	Roadside Assistance: (877) 798-3752|
15	Flagstaff Supercharger	2650 South Beulah Blvd	Flagstaff, AZ 86001	Roadside Assistance: (877) 798-3752|
16	Gila Bend Supercharger	826 W Pima St.	Gila Bend, AZ 85337	Roadside Assistance: (877) 798-3752|

Figure 12.50– Green on the go –Cross Tab Output

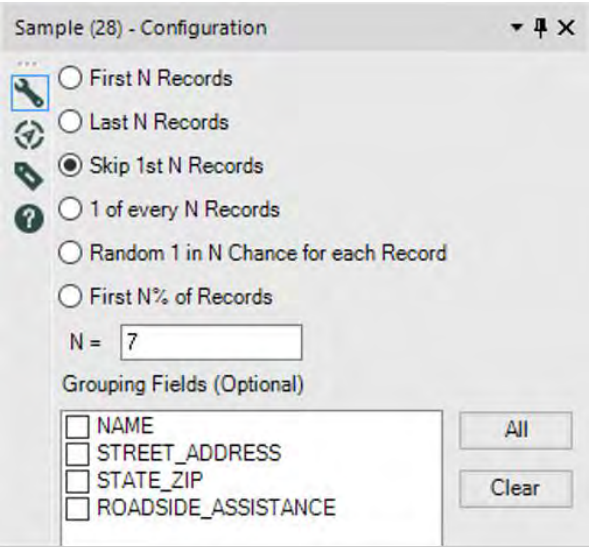


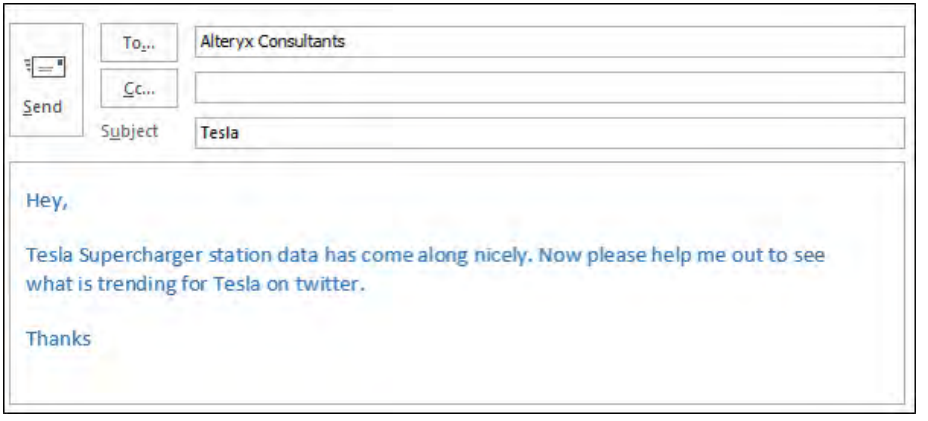
Figure 12.51– Green on the go –Sample Configuration

That's it. We have the data in a required format for the spatial and geo analysis.

Below is the full view of the workflow that we just created.

Figure 12.52– Green on the go –data stream on completion

12.18 What is trending for Tesla?



We would like to search for the hashtag Tesla in Twitter to see what is trending for Tesla. It could be about their car models, new release, supercharger stations, car recalls etc.

Steps below will help us accomplish the goals above,

The Twitter Search tool has to be downloaded from the Alteryx gallery.

To configure Twitter Search, first, we need a Twitter account and have to generate token.

Log in to your Twitter account at https://apps.twitter.com, click "Create a new application", and complete the form (a placeholder website may be used and there is no need for a Callback URL). Once you have submitted the application form, you will be provided with a Consumer Key and Consumer Secret that you can use to configure the tool.

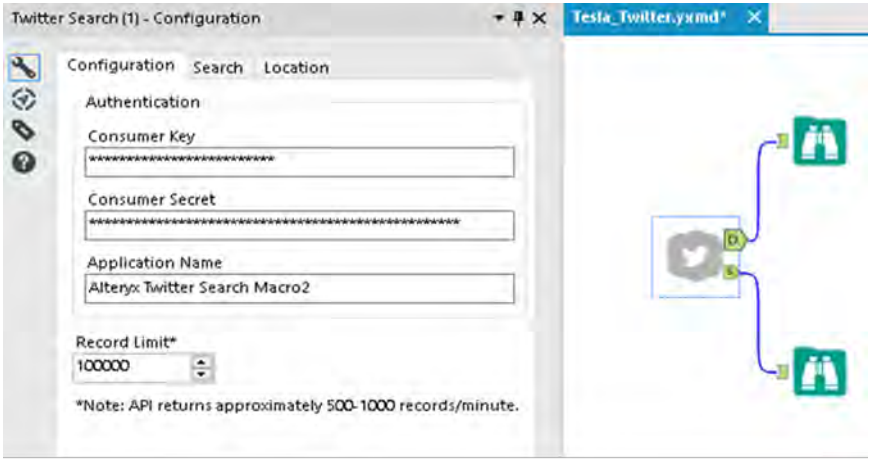


Figure 12.53– What is trending for Tesla –Twitter Search Configuration

Enter the hashtag to look for on Twitter. For us, it is "Tesla"

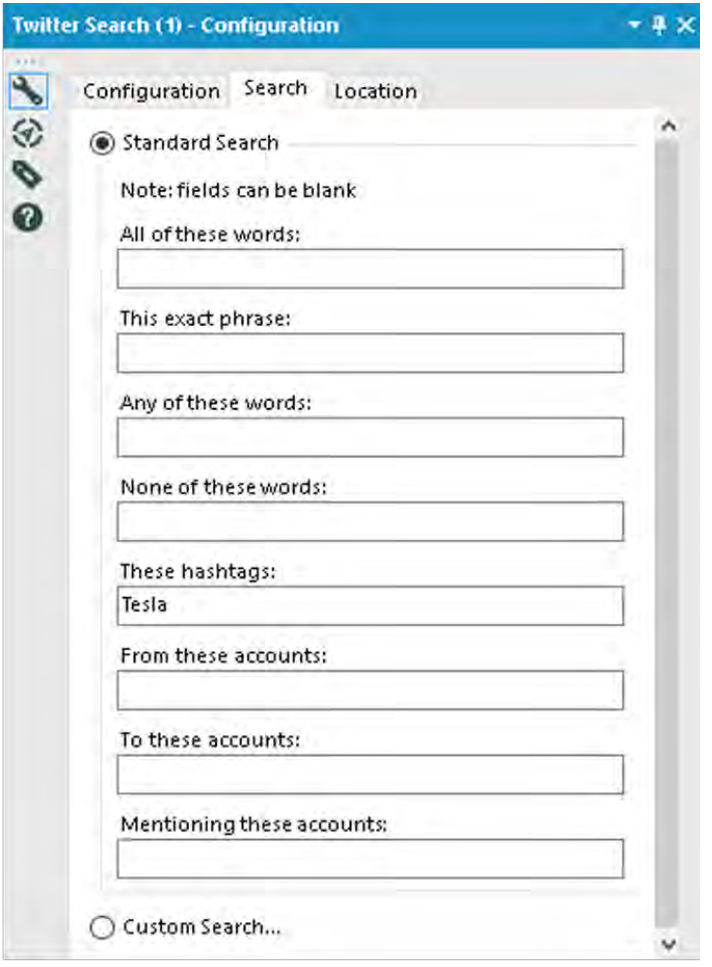


Figure 12.54– What is trending for Tesla –Twitter Search Configuration

We get a lot of information from the twitter. To reduce the size of data, we will select only specific fields that we need for analysis.

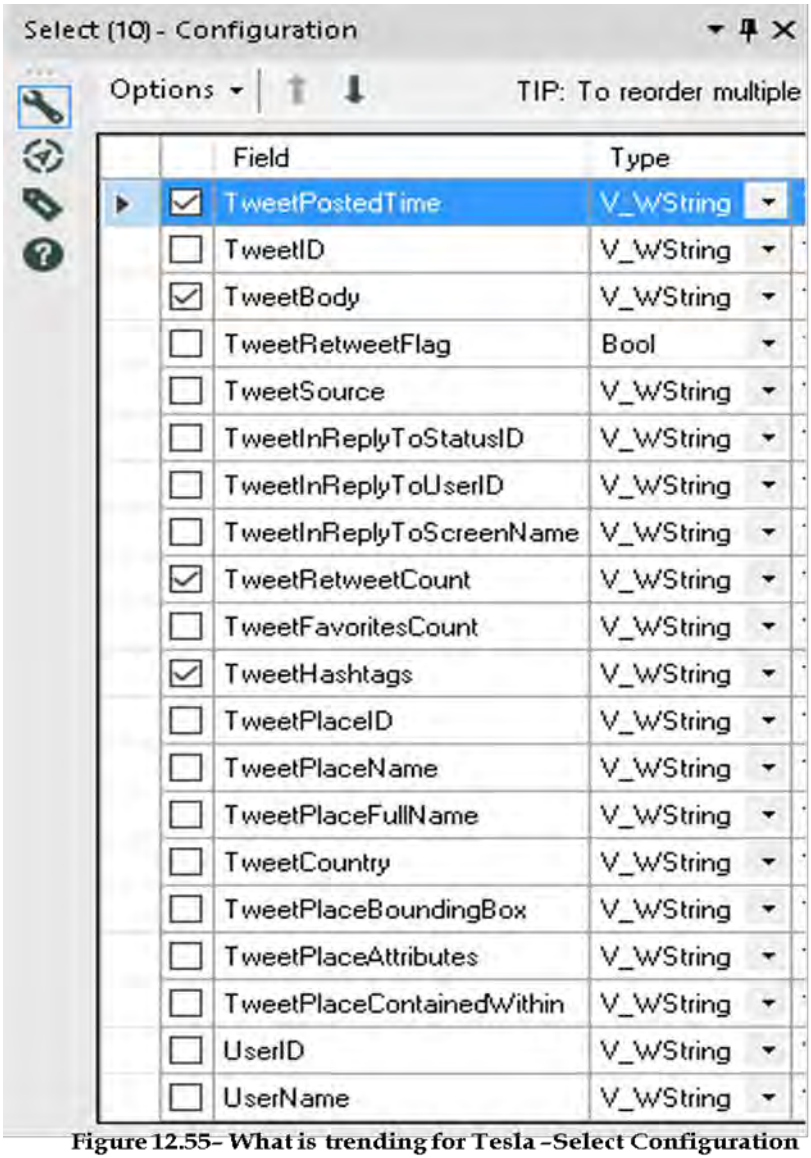


Figure 12.55– What is trending for Tesla –Select Configuration

Let us select below data point from the Twitter search output,

TweetPostedTime
TweetBody
TweetRetweetCount

TweetHashtags
Dynamic or Unknown Fields

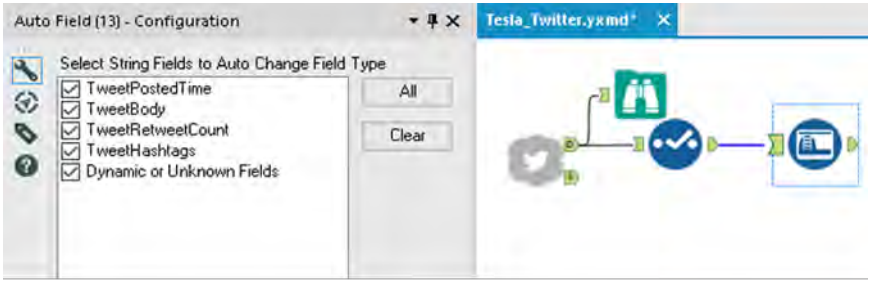


Figure 12.56– What is trending for Tesla –Auto field Configuration

All tweet fields are a string. we use Auto field tool which reads through all the records of an input and sets the field type to the smallest possible size relative to the data contained within the column.

Now we have to find the patterns for these tweets. You can use multiple tools to find the patterns. Like formula toll and Regex tool. In our approach, we will use the Formula tool.

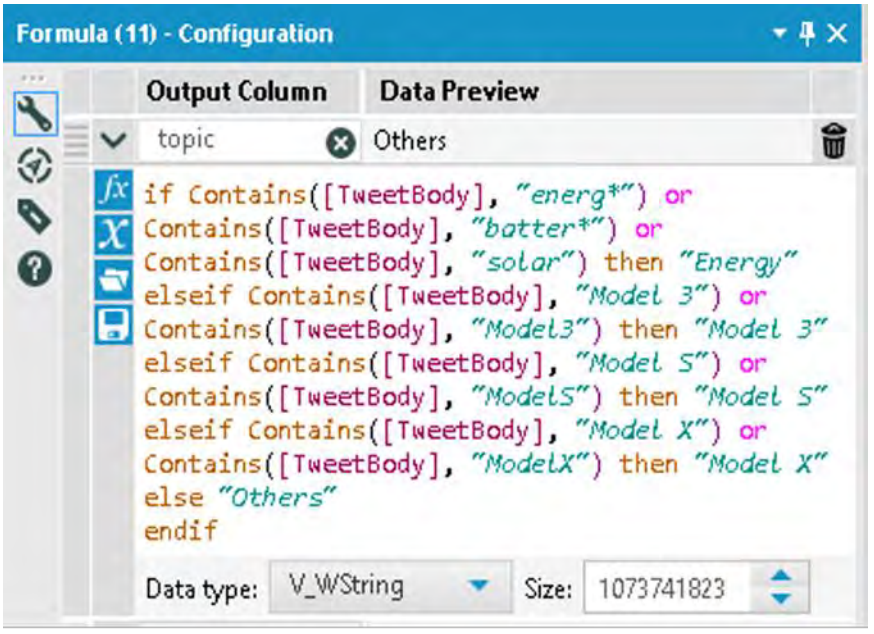


Figure 12.57– What is trending for Tesla –Auto field Configuration

Now we have a lot of rows with the same patterns and we need to aggregate them using the Summarize tool.

We use the Summarize tool twice.

First, to group by topic and get the number of tweets and retweets.

Second, to calculate a total number of tweets. This number or measure will help us in calculating % of a particular topic in all tweets.

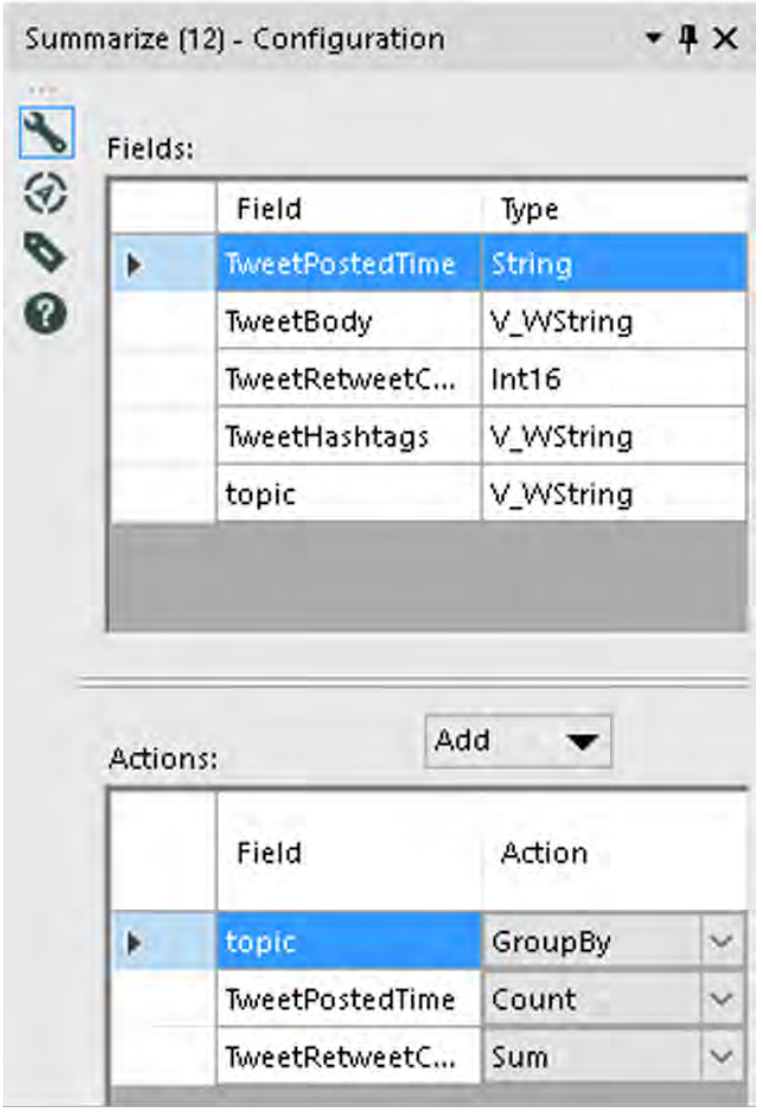


Figure 12.58– What is trending for Tesla –Summarize Configuration

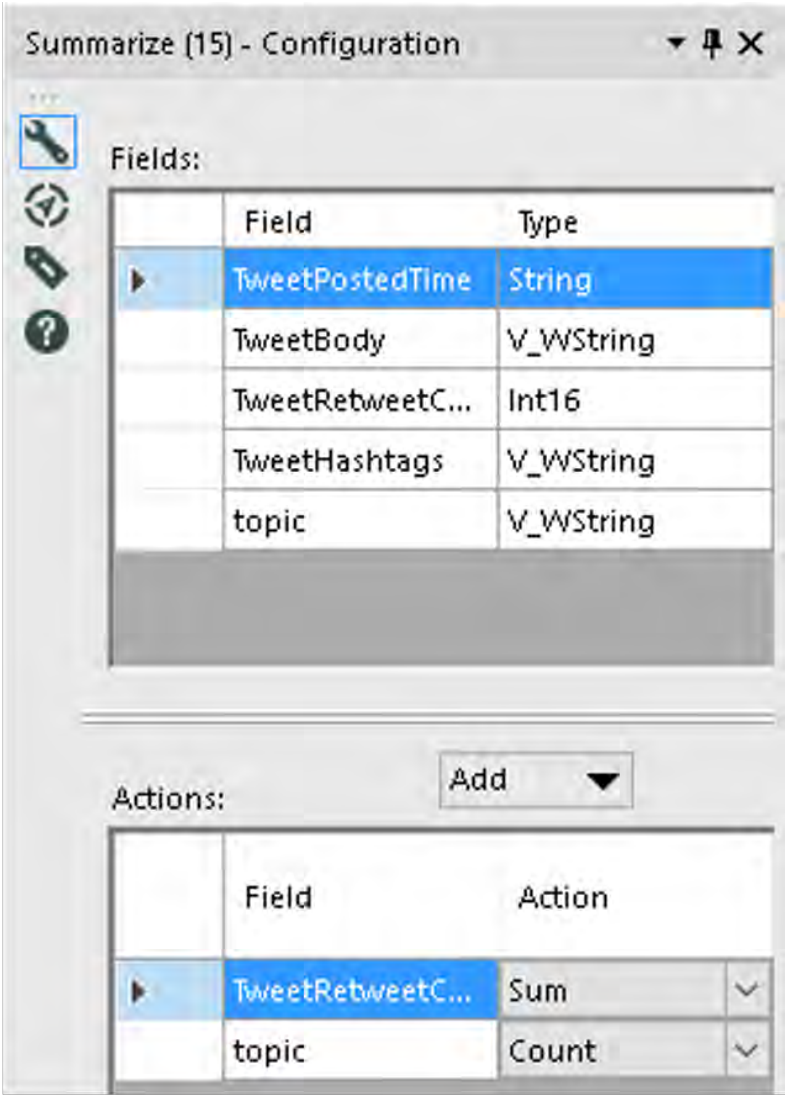


Figure 12.59– What is trending for Tesla –Summarize Configuration

We will use the append fields tools to append a total number of tweets to each topic group.

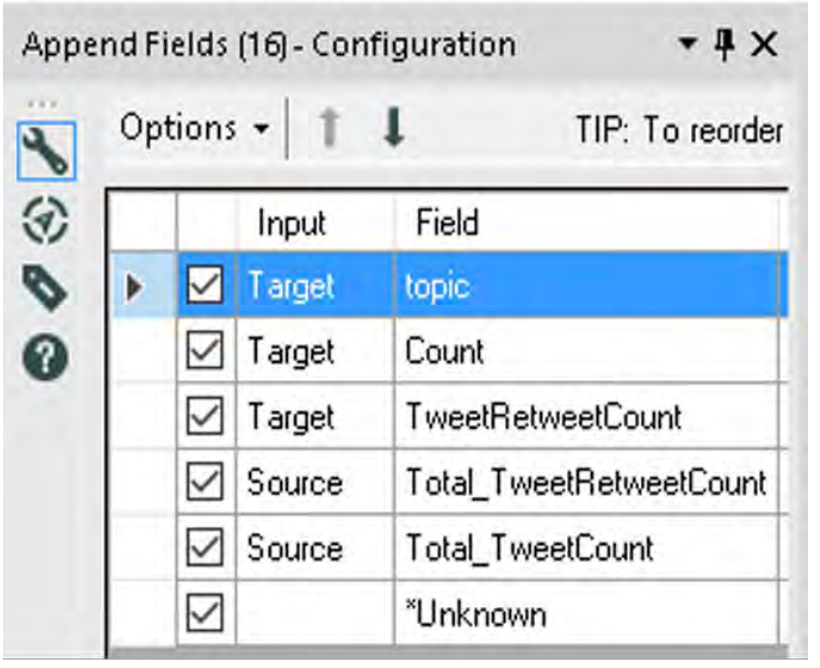


Figure 12.60– What is trending for Tesla –Summarize Configuration

Now Let's check the result. See if we have all the numbers and fields that we need for analysis.

5 records displayed, 5 fields, , 1682 bytes

Table Profile

5 of 5 Fields Cell Viewer

Record #	topic	Count	TweetRetweetCount	Total_TweetRetweetCount	Total_TweetCount
1	Energy	927	7258	893081	17300
2	Model 3	1185	6712	893081	17300
3	Model S	782	109254	893081	17300
4	Model X	229	568	893081	17300
5	Others	14177	769289	893081	17300

Figure 12.61– What is trending for Tesla –Browse tool

Let's calculate % of tweets and retweets for each group.

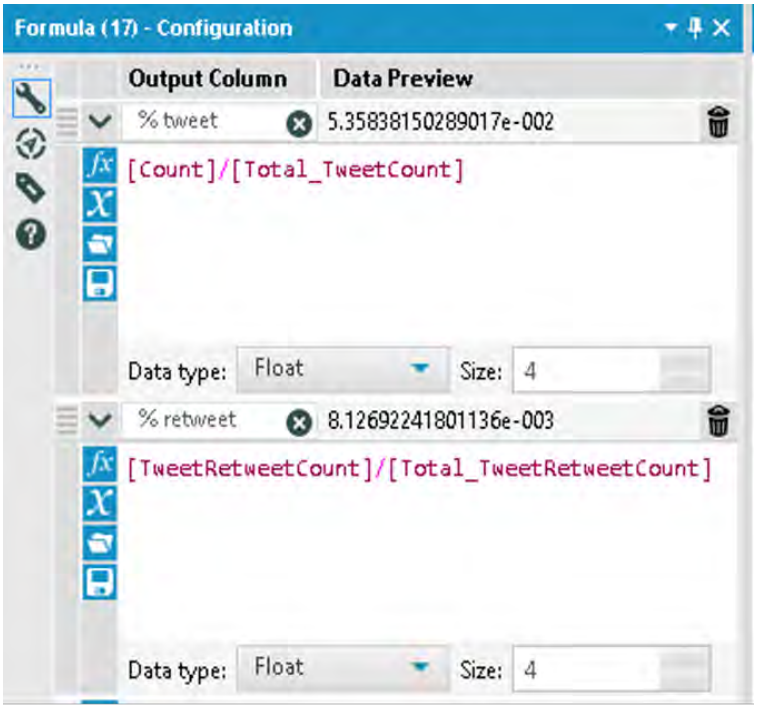


Figure 12.62– What is trending for Tesla –Formula tool

We are almost there. Now let's select all necessary fields and give them appropriate names.

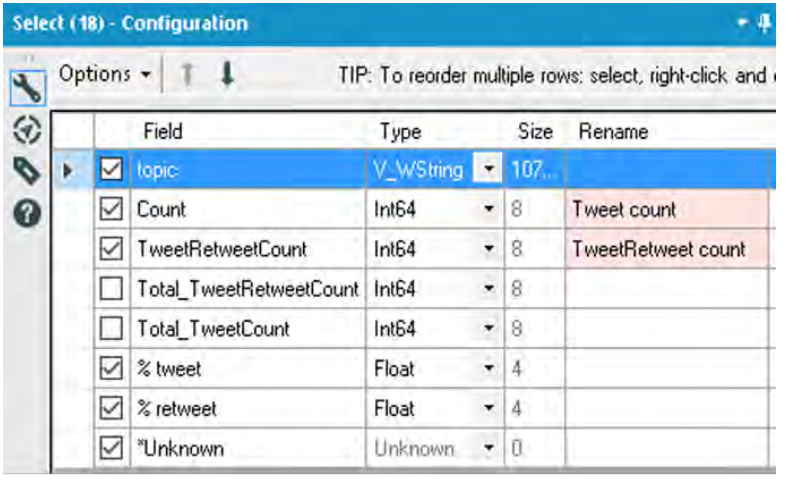
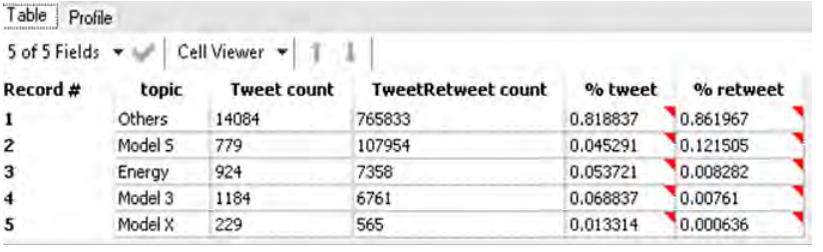


Figure 12.63– What is trending for Tesla –Select tool

Here is our final output. We have sorted the output based on the % of tweets.



Record #	topic	Tweet count	TweetRetweet count	% tweet	% retweet
1	Others	14084	765833	0.818837	0.861967
2	Model S	779	107954	0.045291	0.121505
3	Energy	924	7358	0.053721	0.008282
4	Model 3	1184	6761	0.068837	0.00761
5	Model X	229	565	0.013314	0.000636

Figure 12.64– What is trending for Tesla –Browse the result

The whole workflow is here.

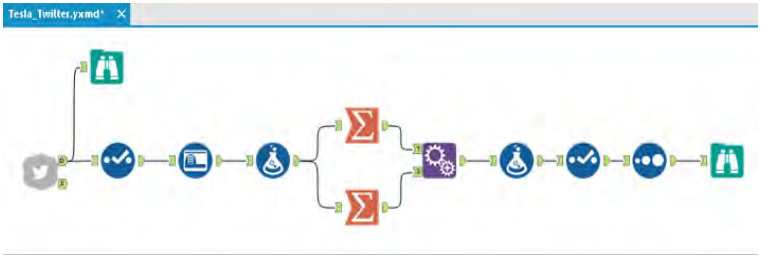


Figure 12.65– What is trending for Tesla –data stream on completion

Authored and Published by

USEREADY

USEReady's mission is to help organizations succeed by fast tracking their business performance. We relentlessly strive towards value-driven innovation and digital transformation using advanced business intelligence, data management, and infrastructure management services. Our global delivery orientation is enriched with winning strategies, rich industrial and technology expertise and customer fanaticism. Our proven track record of success in versatile domains such as capital markets, insurance, healthcare, pharma, retail, media etc. is a testimony to our performance.


Abhilasha Shinde


Anirudh Kalva


Ashish Tergaonkar


Honey Shahi

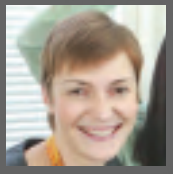

Liyun Zhou


Lokesh Bansal


Lokesh Bhat

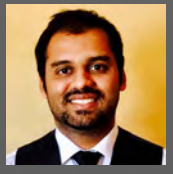

Madhusudhan G


Nishant Thakar


Olga Chirkova


Prateek Kale


Priya Raghuveer


Raghavendra Pai K


Rahul Shetty

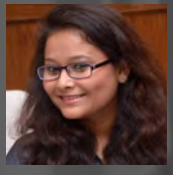

Rahul Upadhye

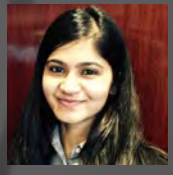

Sachet Kashyap

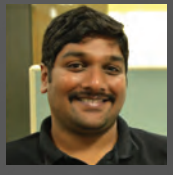

Shamalee Thakur


Smruti Dash


Subrat Das


Suman Joshi


Tanvi Shah


Tharun Kumar


Vijay Gowtham


Vishnudas P N

ISBN 978-1640088085

www.useready.com

contact@useready.com