PROGRAM BOOK FOR

# SHORT – TERM INTERNSHIP

(ONLINE)

Name of the Student : Badarinadh Bonam

Name of the Course : Data Analytics Process Automation

Name of the College : Vishnu Institute of Technology

Registration Number : 21PA1A5415

Period Of Internship : 10 weeks

Name and Address of
Intern Organization : EduSkills Foundation(AICTE) and Alteryx Sparked

## Vishnu Institute of Technology

Bhimavaram, WGDT, Andhra Pradesh

VISHNU
UNIVERSAL LEARNING

MAY - JULY 2023

# DECLARATION

I, Badarinadh Bonam, a student of 3 Year of Bachelor of Technology Program, Reg. No. 21PA1A5415 of the Department of Artificial Intelligence and Data Science, Vishnu Institute of Technology College do hereby declare that I have completed the mandatory internship from May 2023 to July 2023 in Data Analytics Process Automation from EduSkills Foundation under AICTE and Alteryx Sparked under the Faculty Guideship of Mr. Sita Rama Murthy Sir, Department of Artificial Intelligence and Data Science, Vishnu Institute of Technology.

Signature and date

# CERTIFICATION

This is to certify that the Summer Intern Project report submitted by Badarinadh Bonam on the title “Data Analytics Process Automation Virtual Internship” is a record of the summer intern project work done by him during the academic year 2022-2023 in partial fulfillment of Bachelor of Technology.

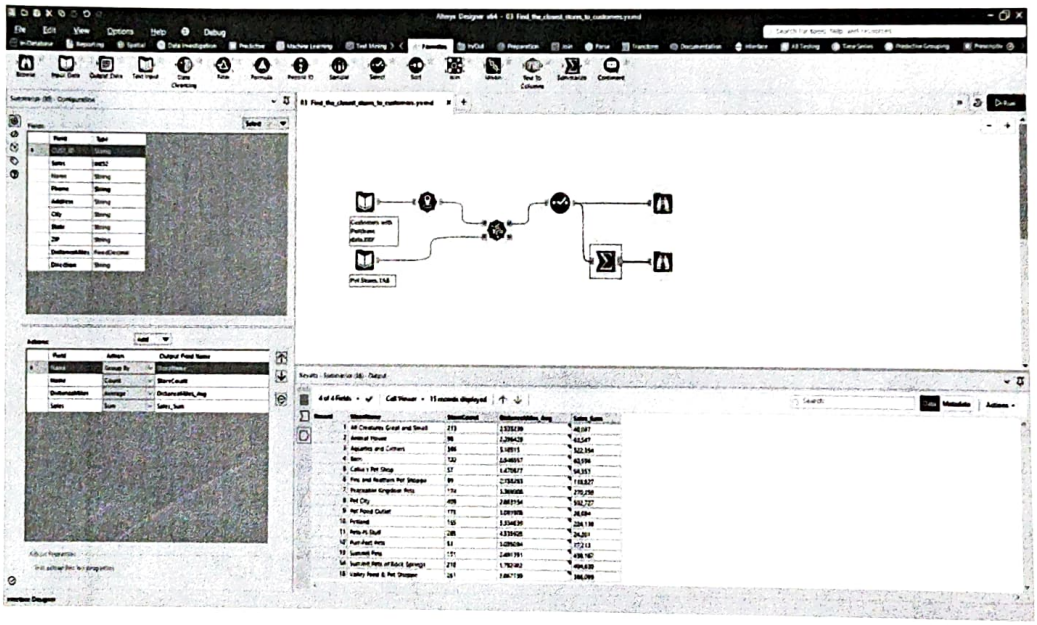
Internal Examiner External Examiner Head of the Department

# INDEX

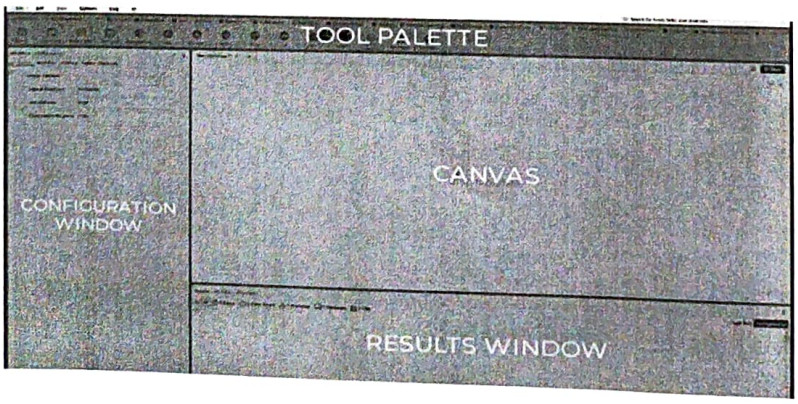
## INTRODUCTION

Alteryx Designer is a windows software application that gives you an intuitive, fun and easy to use drag-and-drop user interface in order to create repeatable workflow processes for analyzing, blending data, and performing advanced analytics (such as predictive, spatial, and prescriptive). You can drag a tool from a tool palette onto the canvas, connect those tools to each other in a process flow that results in one of 3 results: A workflow, an analytic app, or a macro. You can also use these processes you create to quickly and automatically produce results that can be easily shared with others.

Alteryx Designer user interface shows the easy drag and drop interface that users have come to love. Workflows are built with ease but also can support highly complex and complicated processes as well.



The Alteryx Designer contains four primary components that are used to construct a workflow: the Tool Palette, the Canvas, the Configuration window, and the Results window.



**Tool Palette:**

The Tool Palette contains the tools that are used to transform your data. Tools are separated into categories based on the types of functions they perform, such as inputting data, parsing information or creating reports.

Each category's tools share a color and shape, making it easy to visually identify the types of functions being applied in a workflow. If you are unsure as to the function of a given tool, click the tool icon for a description.

For more detailed information on a tool's application or configuration, right click on the tool icon and select "Help" to access the Help Documentation for that specific tool.

**Canvas:**

To create a workflow, simply start adding tools to the canvas. Typically, your first tool is an Input Data tool, which connects to your data source. To relay data from tool to tool, connect the output anchor of one tool to the input anchor of another.

Linking tools together applies a sequence of operations to a data stream. You can add tools to the Canvas in a couple of different ways. First, you can drag a tool from the Tool Palette onto the Canvas. Or simply double click a tool in the Tool Palette.

You can also Search for a tool and drag the tool from the Search results onto the Canvas. In addition, tools can be inserted by right-clicking on the Canvas and selecting the tool from the list of tool categories.

Connect tools by linking their output and input anchors, or just drop a tool in the proximity of another to establish a connection.

**Configure Tools:**

Tools must be configured to take an action on your data. The Configuration window lets you customize that action using tool-specific options. The available configuration options change, depending on the tool that is selected.

Tools that are configured incorrectly will display a red exclamation mark on the tool icon in the Canvas to alert you that the configuration must be fixed to successfully run the workflow.

To change a configuration, select the tool, apply changes in the configuration window and click back to the Canvas to resolve the issue.

**Running a Workflow:**

Running a workflow initiates the flow of data from the input source through the workflow, transforming the data according to the configurations set in each of the tools.

If you're worried about altering your original data, don't be! Your input data remains unchanged as the data moves through the workflow.

Generally, the resulting data is written to a new file or cloud location using an Output Data Tool – although it CAN be configured to overwrite your source data if your particular process requires it.

## CHAPTER - I

## DATA TYPES

In Designer, a column of data is categorized as one of five types: String, Numeric, DateTime, Boolean, or Spatial. In this chapter we'll explore these types in the context of data from an input we will use in our workflow: a survey of trees in New York City. Click a data type to learn more about it.Then, connect the output anchor of the Input Data tool to the input anchor of the Union tool.

**STRINGS:**


A string is a sequence of characters, such as letters, numbers, symbols, and spaces.

The name of a tree, for example, is a value that should be categorized as a string. Even something like a tree's ID, which contains only numbers, is a value that should be assigned a string data type since it's a qualitative attribute.

Qualitative values are those that describe qualities of an object that can be categorized, not measured. These values should not be used for mathematical computations. The length, or the count of characters, of string values can vary.

In Designer, there are two string subtypes to consider: variable & fixed. For data that varies in length, like the names of trees, variable string types should be used. If all values are expected to be the same length, like a six (6) digit Tree ID code, then fixed is a more efficient subtype.

**NUMBERS:**

Numeric data is quantitative, meaning that it can be measured.


Values such as the height of a tree or count of trees in a neighborhood are examples of this data type.

Designer recognizes seven (7) numeric data types that include

integers, or whole numbers, and decimals.

Bytes and integers accommodate whole numbers of different sizes, spanning from numbers containing values from zero (0) to 255, all the way to nine quintillion.

Decimal types (float, double and fixed) allow for additional precision in the form of decimal values.

Float and double are more exacting than fixed, with double being the most precise option, whereas fixed decimals are configured to fit your needs.

**DATETIME:**

Date Time data types can be thought of as three types: Dates, Times and the combination of the two.


Values of this type are familiar and easy to identify but can be formatted many ways.

Datetime data types are important for correctly performing calculations such as "How many trees were planted in the past six months?" or even "What day was it 15,000 days ago?"

**BOOLEAN:**

Boolean data types are the smallest data types in Designer. This data type only has two logical values: 0, or False, and not 0, or True.


It is popular to use this data type when just wanting to categorize a value as one of two values, such as a tree that is "Dead" or "Alive".

**SPATIAL:**

A spatial data type is made up of either a point represented by an X and Y coordinate, a line, or a polygon.

**Viewing Data Types:**

While you may not know the data types in your inputs before building your workflow, Designer provides a window into this information, so you can easily view the data types with which you are working at any point in your workflow.

After running a workflow, simply select a tool to view its metadata in the Results window.

**Viewing Metadata**:

Click the Input Data tool to view the data types associated with the data at the beginning of the workflow.

Then, click the "Metadata" button in the Results Window.

Rather than the actual data values, you now see the metadata associated with the values in each column of our input. Metadata is simply data about data, such as its data type, size,

# CHAPTER - II

# IN/OUT TOOLS

## 1. Browse Tool:

The Browse tool displays data from a connected tool as well as data profile information, maps, reporting snippets, and behavior analysis information in the data.

### View Data Table:

With a Browse tool, the Results window displays all data from the connected tool. With other tools, the Results window only displays up to 1 MB of data. With a Browse tool, the Results window also displays data quality in a bar above each column.
To see the difference in the number of records displayed in the Results window, click the Input Data tool output anchor, then click the Browse tool.

### View Profile Data:

In the Configuration window, the Profile tab displays a chart and metadata for the selected column.
Different types of charts display in the Configuration window depending on the type of data in the column selected in the Results window.
Select columns in the Results window to view its data profile information.

### View Report Data:

In the Configuration window, the Report tab displays report snippets, which are created by tools in the Reporting category.
To view multiple tabs simultaneously, click the View icons in the Configuration window.

### View Spatial Data:

In the Configuration window, the Map tab displays a map of the data. A Browse tool is required to view spatial objects such as points, lines, or polygons.
To change the base map, click Base Map and select a map. Set a default base map in User Settings > Edit User Settings > Dataset Defaults > Reference Base Map.

### More Info

The Browse tool requires temporary space and memory on your computer which may impact performance at run-time. When workflow development is complete, Browse tools can be disabled to improve performance.

Click the canvas to display the Workflow Configuration window. In the Configuration window, click Runtime. Select Disable All Browse Tools. The Browse tools in the workflow are grayed out when disabled.

## 2. Date Time Now Tool:

Use Date Time Now to return a single record: the Date and Time at the workflow runtime, and convert the value into the string format of your choosing. This is a useful tool to add a date-time header for a report

### Configure the Tool

1. Specify your DateTime Language: Select the dropdown to choose the language of the output. The language that you select determines the options available in the Output Format dropdown.
2. Output Format: Select the dropdown to choose the string format for the date-time value returned. The available output format options depend on the selected language.

    - HH:mm:ss
    - day, dd Month, yyyy
    - dd-MM-yy
    - dd-MM-yyyy
    - dd-Mon.-yy
    - dd Month, yyyy
    - dd/MM/yy
    - dd/MM/yyyy
    - dy., Month dd, yyyy
    - MM-dd-yy
    - MM-dd-yyyy
    - MM/dd/yy
    - MM/dd/yyyy
    - Mon dd
    - Month dd, yyyy
    - Month, yyyy
    - yyyy-MM-dd
    - yyyyMMdd
    - yyyy-MM-dd hh:mm:ss
    - MM/dd/yyyy hh:mm:ss

The record Limit is set to 50 to speed up performance during development, especially when the Input Data tool is pulling records from a database.

### Excel File - Define a Data Range

After an excel file is selected, a named range may be selected. If the excel file contains a named range, the dropdown defaults to the available range.
If no named range exists, sheet names are selectable in the dropdown.

### Excel File - List of Sheet Names

Instead of inputting rows of data, when connected to an excel file, a list of the names of the sheets contained in that file can be read in.

### Comma-Delimited Text File (.CSV)

The default options are usually appropriate when reading in a . CSV file.
Note that the default comma delimiter can be changed via the Delimiters option.

### Other file types - Classic Mode

The Input Data tool can connect to many other file types. To see a complete list:

1. Drag an Input Data tool onto the canvas.
2. In the Connect a File or Database dropdown, select File.
3. Click the Files of type dropdown to see the list of supported files.

You can also select an Input Data tool and press F1 to see a list of supported file types.
Database Connections - SQL Server, Oracle, Hadoop - Classic Mode
To connect to these common database types:

1. Drag an Input Data tool onto the canvas.
2. In the Connect a File or Database dropdown, select Microsoft SQL Server, Oracle, or Hadoop.
3. Follow the prompts to connect to your database.

### Database Connections - Other Databases - Classic Mode

The Input Data tool can connect to many database systems. To connect to other database types:

1. Drag an Input Data tool onto the canvas.
2. In the Connect a File or Database dropdown, select Other Databases.
3. Select a database type and follow the prompts.

For a complete list of supported databases, and to learn how to connect to various database types, select an Input Data tool and press F1.

This Output Data tool is configured to create a separate. CSV file for every two rows in the data stream using the Max Records Per File option.

Note: The pipe ('|') is set as the delimiter rather than the default, comma.

3 . CSV files are generated:
1) "OutputToolExample_WithMaxRecords.csv"
2) "OutputToolExample_WithMaxRecords_1.csv"
3) "OutputToolExample_WithMaxRecords_2.csv"

**Write-Output to Another File Type, a Database, or Saved Data Connection**
The Output Data tool can write to many file and database types. You can also write to a saved data connection or a connection shared through a Private Gallery when using Alteryx Server.

1. Drag an Output Data tool onto the canvas and connect the Text Input tool to it.
2. In the Write to File or Database dropdown, select an option:

File - Select the Save as type dropdown to see the list of supported files.
Microsoft SQL Server - Follow the prompts to connect to your database.
Oracle - Follow the prompts to connect to your database.
Hadoop - Follow the prompts to connect to your database.
Other Databases - Select a database type and follow the prompts.
Saved Data Connections - Select an existing connection or Add a Gallery.

To see a list of supported file types and databases, select an Output Data tool and press F1.

## 5. Text Input Tool:

The Text Input tool allows you to create a stream of data inside a workflow without a dependency on a separate file or database. The data set becomes part of the workflow.

**Manually Typed Data**
Add text to a Text Input tool by clicking each column (or column header) and typing in the desired data.
View the Results window and click the Metadata button to view the data types of the data in each column.

**More Info**

# PREPARATION TOOLS

1. **Data Cleansing Tool:**

The Data Cleansing tool performs basic data cleansing operations such as replacing null values, removing punctuation, and modifying capitalization.

**Replace Nulls, Numeric Columns**
This configuration replaces null values in the Score column with zeros. This option can only be used for numeric values.

**Fix Common Problems in Text Columns**
This configuration replaces null values with blank spaces and removes leading and trailing spaces in the selected columns containing text. Compare input and output data, by clicking the Data Cleansing tool input and output anchors, or by clicking the input and output anchor icons located on the left side of the Results window.

**Remove Unwanted Characters, Text Columns**
This configuration removes extra formatting characters, words, or other punctuation from the selected column.

**Standardize Capitalization**
This configuration changes the capitalization of letters and removes leading and trailing whitespaces from the selected column.

**More Info:**
The Data Cleansing tool is a macro. You can customize a macro to address your specific needs. Right-click the tool and select Open Macro. Review the workflow and revise tool configurations to meet your needs.

2. **Filter Tool:**

The Filter tool splits a data stream into 2 streams based on a conditional expression. Rows that satisfy the condition flow out of the True anchor; the rest come out of the False anchor.

### Basic Filter, Greater Than

The Basic filter option allows for easy comparisons of columns against a static value.
Records with CustomerID greater than 30 are True; the rest are False.

### Basic Filter, Is Not Null

Different operators are displayed in the dropdown list based on the selected column's data type.
The LastName column, type string, is filtered against Null values. Galileo's LastName is Null, so his row is False.

### Basic Filter, Date Expressions

Columns that have a data type of date or date/time allow users to dynamically filter on Today, Tomorrow, and Yesterday; you can also filter by selecting a Fixed Date.
Rows that are less than or equal to Tomorrow's date are output as True.

**Basic Filter, Date, and Period Expressions**

The operator's Date range, Start date and periods after, and End date and periods before requiring date selection. Operators that require a single fixed date also require a period to be selected.
Rows that are within a period of 2 days from Today's date are True.

### Custom Filter, Multiple Columns

The Custom Filter option allows for querying against multiple columns.
Rows with a JoinDate greater than or equal to the FirstPurchaseDate are True.

### Custom Filter, Complex Expression

The Custom Filter enables the combining of functions to create more complex expressions.
Rows from the South region or from a region with the word West in it are True.

### More Info:

To reference a column or constant in a Custom filter expression, select the blue Columns and Constants "X" button, or type a left bracket "[" in the expression box and choose from the list that appears.

If a column has DateTime data, a check box displays the option to filter using just Date or to filter using Date and Time data.

3. **Formula Tool:**

The Formula tool enables you to perform a variety of calculations and operations to create new data columns or update existing columns.

**Create a New Static Column**
RecordSource is added under the Output Column to create a new column with a text entry showing the source of the records.

**IF/ELSE Conditional Formula**
A new Region column is added based on the values in the Latitude column using the IF/Then/Else function. To search for functions, click the Functions "fx" button to the left of the expression box or begin typing and choose from the list that appears.

**New Modified Column**
A new column is created based on the City column, with a function to set consistent capitalization.

**Modify Existing Column**
Instead of adding a new Output Column, an existing column (City) is chosen and the capitalization function is applied directly.

**Multiple Columns**
A single Formula tool can contain multiple formulas for multiple columns. Here, three of the examples above are represented in one tool.

**Complex Formula, Single Step**
A new column, AverageSpendPerVisit, is created in a single expression by dividing two existing columns and rounding the result to the nearest whole number.

**Complex Formula, Multiple Steps**
In this Formula tool, the same calculation as above is performed but in 2 steps:
1) Spend is divided by the number of Visits.
2) The result of the expression above is then rounded.

**More Info:**

The Sort tool arranges rows in a table in an ascending or descending alphanumeric order based on the values of one or more specified data columns.

**Sort by 1 Column**
Rows are sorted per the "CustomerID" in ascending order.

**Sort by 2 Columns**
LastName is sorted in ascending order first and then the same LastName values are sorted by FirstName.

Note how the Null value in the LastName column comes first in the sort.

The Sort tool can be used to sort as many columns as needed.

**Sort by a Numeric String Value**
The score (a string field containing numbers) is sorted in ascending order.

The value 22 is placed before 4 because numeric strings are sorted left to right, character by character.
Sort with Use Dictionary Order Selected
The score is sorted in ascending order and Use Dictionary Order is checked.

Unlike above, the value 22 is now placed after 4 because using Dictionary Order sorts numeric strings from the smallest number to the largest.

7. **Unique Tool:**

Use Unique to distinguish whether a data record is unique or a duplicate by grouping on one or more specified fields, then sorting on those fields.

**Configure the Tool**
**Column Names:** Select the columns where you want to find unique values.

- Use the Select All button to compare entire records. The data is sorted based on the Unique columns. Therefore if you want a specific sort order, use the Sort tool to assign the specific sort order of the file prior to using

# CHAPTER - IV

# JOIN TOOLS

## 1. Append Fields Tool:

Use Append Fields to append the fields of one small input (Source) to every record of another larger input (Target). The result is a Cartesian join. In a Cartesian join, every row from one table is joined to every row of another table. For example, if table A has 100 rows, and table B has 1,000 rows, the Cartesian join of these two tables results in 100,000 rows.

### Connect Inputs

The Append Fields tool accepts 2 inputs:

**T anchor (Target):** The larger data stream that records are appended to.
**S anchor (Source):** The smaller data stream that provides the records that are added to the Target stream

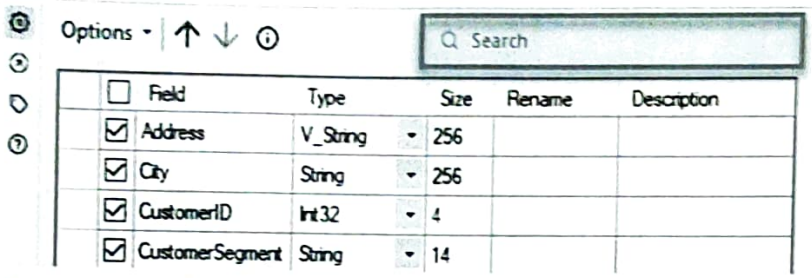
### Configure the Tool

Use the table in the Configuration window to modify the incoming data stream. Each row in the table represents a column in your data. The Field column in the table identifies the name of each column in the data and auto-sizes to fit column (field) names without cutting off any text (up to 40 characters)

### Search Fields

Before you start updating your fields (columns), you might want to limit your list so that you can perform updates on only a subset of the fields. This is also really helpful if your dataset contains many fields.

To do this, you can use the Search box at the top of the Configuration window. Enter a keyword and the Append Fields tool searches the Field, Rename, and Description columns to return matches. The search is not case-sensitive.



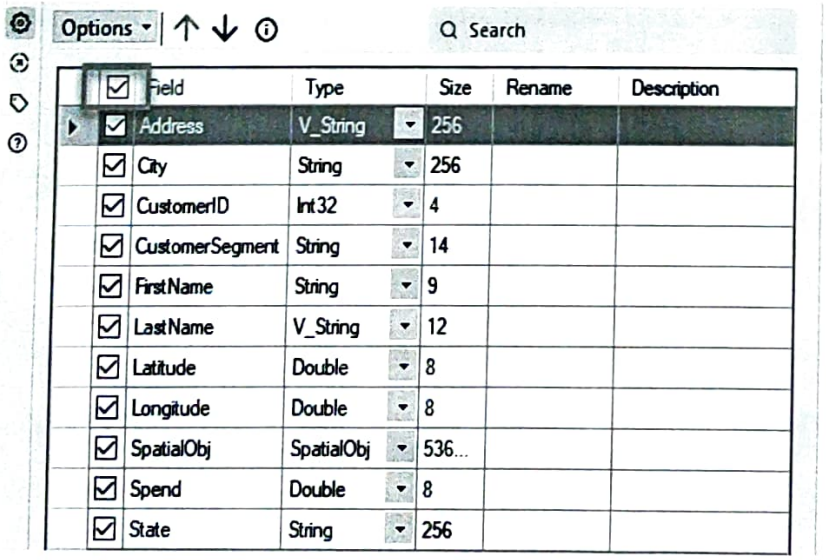

You can then perform various actions (like select, deselect, rename, etc.) on only the fields that were returned via your search. Think of the Search box as a way to filter your list of fields so that you can update only a subset of your data columns.

To view your entire list of fields again, use the "x" icon to clear out the Search box.

### Select, Deselect, Sort, and Reorder Columns

### Select and Deselect Fields/Columns

To include a column in the dataset, check the check box to the left of the column name. Uncheck the check box to exclude the column. You can also use the select and deselect all check box at the top of the table to quickly select and deselect all visible fields.




### Sort Columns:

To sort the columns of data based on the column name...

- Click on the column name to sort in ascending order.
- Click on the column name a second time to sort in descending order.

- **Remove Prefix or Suffix:** Remove the prefix or suffix from the selected or highlighted column name.
- **Clear All Renames:** Remove the new name for all columns.
- **Clear Highlighted Renames:** Remove the new name for all highlighted columns.
- **Revert All to Original Type & Size:** Undo all changes to type and size in all columns and use the original values.
- **Revert Highlighted to Original Type & Size:** Undo changes to type and size in the selected or highlighted columns and use the original values.
- **Forget All Missing Fields:** Remove all columns that are no longer included in the data.
- **Forget Highlighted Missing Fields:** Remove all highlighted columns that are no longer included in the data.
- **Deselect Duplicate Fields:** Deselect the second column when duplicate column names exist. This option is only available with multiple inputs.

## 2. Find Replace Tool:

Use Find Replace to find a string in one column of a dataset and look up and replace it with the specified value from another dataset. You can also use Find Replace to append columns to a row.

### Tool Components

**The Find Replace tool has 3 anchors:**

### Input anchors:

**F Input anchor:** This input is the initial input table ("F" for "Find"). This is the table that is updated in the tool's results.
**R Input anchor:** This input is the lookup table ("R" for "Replace"). This is the table that contains data used to replace data in (or append data to) the initial input.
**Output anchor:** The output anchor displays the results of the Find Replace tool.

### Configure the Tool

The Find Replace tool configuration is comprised of 2 sections: Find and Replace.

#### Find Section

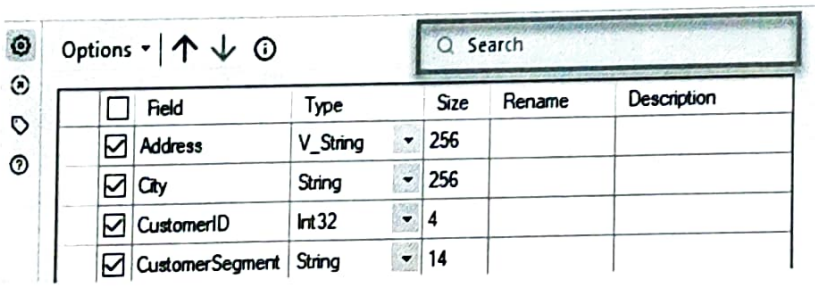
1. Choose the radio button that best describes the part of the field that contains the value to find:

table identifies the name of each column in the data and auto-sizes to fit column (field) names without cutting off any text (up to 40 characters).

### Search Fields

Before you start updating your fields (columns), you might want to limit your list so that you can perform updates on only a subset of the fields. This is also really beneficial if your dataset contains many fields.

To do this, you can use the Search box at the top of the Configuration window. Enter a keyword and the Join Multiple tool searches the Field, Rename, and Description columns to return matches. The search is not case-sensitive.



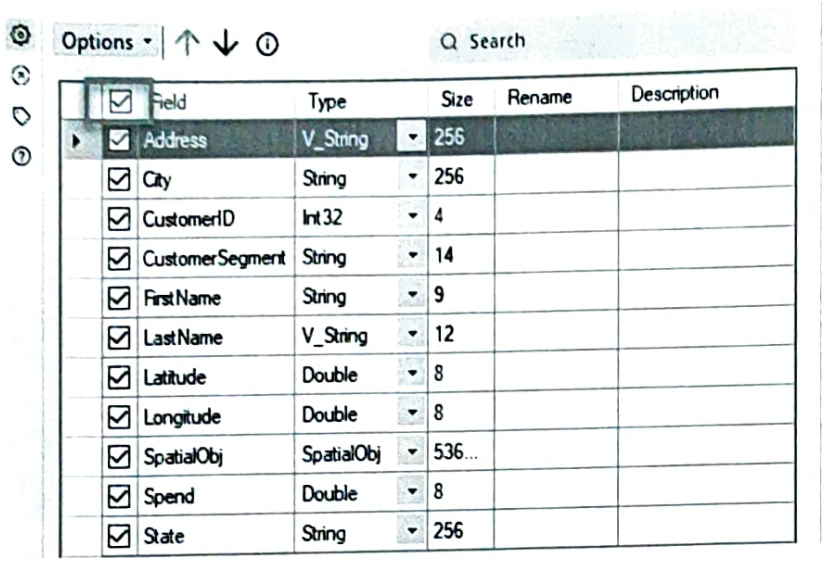

You can then perform various actions (like select, deselect, rename, etc.) on only the fields that were returned via your search. Think of the Search box as a way to filter your list of fields so that you can update only a subset of your data columns.

To view your entire list of fields again, use the "x" icon to clear out the Search box.

### Select, Deselect, Sort, and Reorder Columns

#### Select and Deselect Fields/Columns

To include a column in the dataset, check the check box to the left of the column name. Uncheck the check box to exclude the column. You can also use the select and deselect all check box at the top of the table to quickly select and deselect all visible fields




## Sort Columns

To sort the columns of data based on the column name...

Click on the column name to sort in ascending order.
Click on the column name a second time to sort in descending order.

## Reorder Columns

To reorder the columns of data...

- Select to highlight a row, or select and drag to highlight multiple rows.
- Use the Move Up or Move Down arrows, or right-click and drag, to move the rows to a new location.

The Unknown column is selected by default. It allows new columns in the data. Move the column to the location where you want a new column to be.

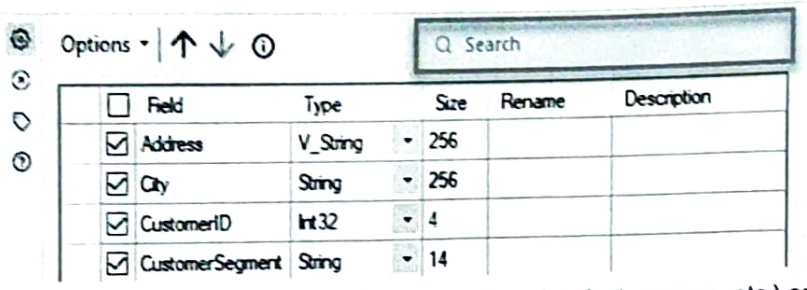
## Modify Data Type and Size

### Data Type

Use the Type dropdown to change the data type of a column in your dataset.

### Data Size

To change the supported length (characters for string and numeric fixed decimal types) or measurement (bytes for other numeric types) of data in a column, select Size and enter a number. Size varies by data type and you can edit it for fixed decimal numeric types and all string types.



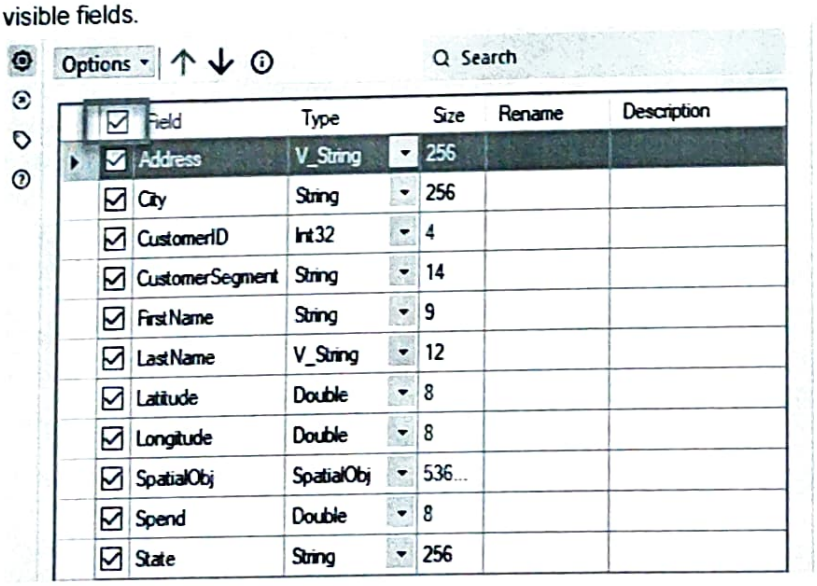

You can then perform various actions (like select, deselect, rename, etc.) on only the fields that were returned via your search. Think of the Search box as a way to filter your list of fields so that you can update only a subset of your data columns.

To view your entire list of fields again, use the "x" icon to clear out the Search box.

**Select, Deselect, Sort, and Reorder Columns**

**Select and Deselect Fields/Columns**

To include a column in the dataset, check the check box to the left of the column name. Uncheck the checkbox to exclude the column. You can also use the select and deselect all checkbox at the top of the table to quickly select and deselect all visible fields.




**Sort Columns:**
To sort the columns of data based on the column name...

- Click on the column name to sort in ascending order.
- Click on the column name a second time to sort in descending order.

## 5. Union Tool:

Use Union to combine 2 or more datasets on column names or positions. In the output, each column contains the rows from each input. You can configure how the columns stack or match up in the output.

**Configure the Tool**
Mode: Choose the configuration mode. The default setting is Auto configure by name.

- Auto configure by name: Stack data by column name.
- Auto configure by position: Stack data by the column order in the stream.

Manually configure columns: Allows you to manually specify how to stack data. When you choose this method, the columns in each input are displayed (indicated by row #1, #2, etc.).

**Properties: Auto Config**

**When Columns Differ**
For the auto-configuration modes, you must select how to handle columns that differ.

**In the first dropdown, choose your error handling option...**

- Error - Stop Processing Records: Throw an error in the Results window and stop processing records.
- Warning - Continue Processing Records: Throw a warning in the Results window, but continue processing records.
- Ignore - Continue Processing Records: Ignore columns that differ and continue processing records.

**In the second dropdown, choose your output option...**

- **Output All Fields:** Output includes all columns. Null values populate empty columns.
- **Output Common Subset of Fields**: Output includes only the columns that each input has in common.

**Properties: Manually Configure Fields**
For the manually configure columns mode, you have to configure your Output Columns in the Properties section.

To begin, your data streams are staggered horizontally and vertically so that the data from each input dataset are in different cells.

1. (Optional) In the top-right dropdown, you can begin by selecting either By Position or By Name. Select Reset to reset the columns. Use this option if you know that your data streams have some columns that match by either position or name.
2. Next, use the arrows to begin stacking your data. Select a cell and select the left arrow or right arrow to stack it with the data field it matches.
3. Select Non Blocking - Metainfo Will Not Change to pass data rows downstream without waiting for all inputs to send data. Do not use this mode if the upstream metainfo will change between configuration time and runtime.

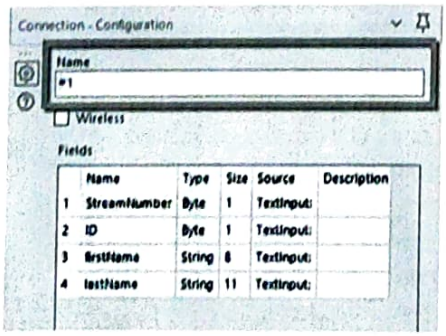
**Output Order**
Under Output Order, check Set a Specific Output Order to specify which input dataset's data displays first in the output dataset. Once checked, select one of the data streams and select the up arrow or down arrow to reorder.

**Understanding the Output**
Two aspects of the Union tool output are important to understand, the column names and the order of the data.

**Understanding Output Data Column Names**
The column names that are used in the output dataset are pulled from the input stream with the first alphabetical/numerical value.

By default, your input data streams are labeled #1 and #2 based on the order you connected them to the Union tool input anchor. So, if the column names differ, the output dataset will use the column names from the #1 input dataset.

If you prefer to use the column names from the #2 input dataset, you can change the Name for the input connections. To do so, select the input streams and enter new values in the Name field of the connections. The output column names are taken from the connection with the first alphabetical/numerical value in its Name.




**Understanding Output Data Order**
The default output order often corresponds to the order you connected your input datasets to the Union tool input anchor but might vary.

# CHAPTER - V
# PARSE

## 1. DateTime Tool:

Use DateTime to transform date-time data to and from a variety of formats, including both expression-friendly and human-readable formats. You can also specify the language of your date-time data.

**Configure the Tool**

Connect a data source to the input anchor of the DateTime tool. In the Configuration window, select the format to convert. There are 2 options:

1. Date/Time format to string: Converts the date-time data format to a string.
   - Select the date/time field to convert: Select the dropdown to choose a date-time field (column). If no date-time columns are available, the field appears dimmed and reads No Suitable Input Fields Available.
   - Specify the new column name: Enter a new column name for the converted data, or use the default DateTime_Out.
   - Specify your DateTime Language: Select the dropdown to choose the language for your new string column output. Go to the Specify Your DateTime Language section below for a list of formats per language.
   - Select the format for the new column: Select the format for your new column from the list, or select Custom to create your own format via the Specify a custom format for the new column field. Refer to the Custom Format section below.

2. String to Date/Time format: Converts string data to a DateTime format.
   - Select the string field to convert: Select the dropdown to choose a string field (column) to convert to a date-time format. If no string columns are available, the field appears dimmed and reads No Suitable Input Fields Available.
   - Specify the new column name: Enter a new column name for the converted data, or use the default DateTime_Out.
   - Specify your DateTime Language: Select the dropdown to identify the language of the incoming string data. Go to the Specify Your DateTime Language section below for a list of formats per language.
   - Select the format that matches the incoming string field: Select the format of the incoming string from the list, or select Custom to identify a different format via the Specify the format of the incoming string field section. Refer to the Custom Format section below.

- yyyyMMdd
- HH:mm:ss
- Custom

## 2. Text to Columns Tool:

Use Text To Columns to take the text in one column and split the string value into separate, multiple columns (or rows), based on a single or multiple delimiters.

**Configure the Tool**

1. Select the Column to split.
2. Delimiters: Enter the delimiters to use to split the data. Each character is treated independently, meaning you can't delimit on a word. To split data on a word, use the RegEx tool. You can use individual characters or the white space characters from this table:

| White Space | White Space Character |
|---|---|
| Tab | \t |
| New Line | \n |
| Space | \s |
| Space or Tab | \s\t |

3. Choose the split method:

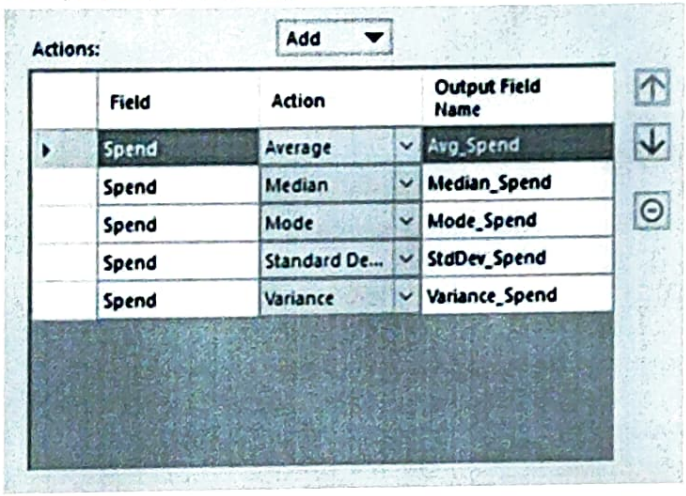
    - Split to columns: Split a single column of data at each instance of the specified delimiter into multiple columns.

    - Number of columns: Set how many columns to create.
    - Extra characters: Let Alteryx know what you want it to do with any extra characters left over. Go to Alteryx Community Tool Mastery Article for tips and tricks including how to find out how many columns your data will be parsed into.
    - Leave extra in the last column: Data that extends past the split is appended to the value in the last column.
    - Drop extra with warning: Data that extends past the split is dropped and a warning is generated indicating that there was excess information.

- Spatial: Only fields with a spatial data type are selected.

2. With your fields selected, select the Add dropdown above the Actions section. There are various actions to choose from. Actions are not available if they are not compatible with the selected data types.
3. Select the action that you want to perform. Your fields are added to the Actions section.

You can perform multiple actions on a single field. To do so, select and add the field once per each action that you want to perform.




**Actions**

The Actions section contains data fields added from the Fields section. Several options are available once fields have been added to the Actions section:

- Reorder: Select a field and use the up or down arrow buttons to change the order of the fields. This will also update the field order of the tool output in the Results window.
- Remove: Select an action field and use the remove button to remove the action from the Actions section.
- Change Action: To change the action, select the dropdown in the Action column and choose a compatible action.
- Rename Output Field: To rename a field, first select the field and then input a new name into the Output Field Name column.

**Action Properties**

Percentile, Concatenate, and most Finance actions require you to specify additional properties.

- When additional properties are required, specify these via the Properties section at the bottom of the tool Configuration window. Go to Summary Actions for information on specific properties.
- If additional properties are not required, the Properties section will not be available.

**Summary Actions**
This list describes the types of summary actions that the Summarize tool can perform:

1. **Summary Actions**

    - Group By: Combines database records with identical values in a specified field into a single record. All of the resulting data from the records in a group are then summarized. Any non-blob or spatial object has this option. If no Group by field is specified, the entire file will be summarized.
    - Sum: Returns the sum value for the group. The sum is calculated by adding all of the values of a group.
    - Count: Returns the count of records in the group.
    - Count Non Null: Identical to Count, except it is only counting those records that are not null. Null means there is no value set for the record. This is different than a zero or an empty string.
    - Count Distinct: Returns the count of unique records in the group.
    - Count Distinct Non Null: Identical to Count Distinct, except it is only counting those records that are not null. Null means there is no value set for this record (different than a zero or an empty string).
    - Count Null: Identical to Count, except it only counts those records that are null. Null means there is no value set for the record. This is different than a zero or an empty string.
    - Min: Returns the minimum value.
    - Max: Returns the maximum value.
    - First: Returns the first record in the group, based upon its record position.
    - Last: Returns the last record in the group, based upon its record position.

2. **Finance:**

**Net Present Value (NPV):** Calculate the net present value of an investment. NPV is a measure of future cash flow over the life of an investment. NPV Parameters:

- Discount Rate (Per Period): The discount rate as a percentage. The default value is 8%.

**Net Present Value w/ Dates (XNPV):** Calculate the Net Present Value for an investment with dates. XNPV Parameters:

- Finance Rate: The finance rate as a percentage. The default value is 8%.

## 2. Transpose Tool

Use Transpose to pivot the orientation of the data table. It transforms the data so you may view Horizontal data fields on a vertical axis. The Transpose tool is the reverse application of the Cross Tab tool, which pivots the orientation of the data from vertical to horizontal.

This tool is useful for extracting non-conforming data. There is no limit to the number of records or fields that can be transposed.

**Configure the Tool**

1. Key Columns: Select the columns (fields) to pivot the table around. This name remains on the Horizontal axis, with its value replicated vertically for each data field selected.
2. Data Columns: Select all the columns to carry through the analysis.
3. Use Select All to check all of the options, or Deselect All to clear all selections.
4. Missing Columns: Select the action that you want to take if columns are missing:
   - Error: Throw an error in the Results window and stop processing records.
   - Warn: Throw a warning in the Results window, but continue processing records.
   - Ignore: Ignore missing columns and continue processing records.

**Key Columns Example**

The Key Columns you select replicate vertically and create a row for each Data Column you select to transpose.

For example, you have a table containing 2 rows and 4 columns. When you transpose the table and select a Key Column without deselecting any Data Columns, you end up with 6 rows and 3 columns along with the selected Key Column, Name (represents the name of the remaining selected Data Columns), and the Value of the selected Data Columns.

If you deselect the Data Columns, the number of rows after the Transpose tool decreases.

The number of rows you end up with after transposing the table is equal to the number of initial rows multiplied by the number of Data Columns selected, minus any columns you selected as Key Columns.

## 3. CROSS TAB TOOL:

Use Cross Tab to pivot the orientation of data in a table by moving vertical data fields onto a horizontal axis and summarizing data where specified. The Cross Tab tool is the reverse

application of the Transpose tool, which pivots the orientation of the data from horizontal to vertical. There is no limit to the number of rows or columns that can be passed through the Cross Tab tool.

**Configure the Tool**

1. A list of all available data values displays in the Group data by these values section. Select the values that should be used to group the data. Data with identical values are grouped together into a single row.
2. Select a value from the Change Column Headers dropdown. A new column is created for each unique value.
3. Select a value from the Values for New Columns dropdown. These values are used to populate the new columns.
4. Select a Method for Aggregating Values for combining multiple values in a field. Available options depend on the data type of the value selected in the Values for New Columns field.
    - Options for String data include...
    - Concatenate: Separates the values using the separator specified in the Separator field.
    - First: Displays the first found value.
    - Last: Displays the last found value.
    - Options for Number data include...
    - Sum: Sums the values.
    - Average: Calculates an average of the values.
    - Count (without Nulls): Counts the number of values excluding null values.
    - Count (with Nulls): Counts the number of values including null values.
    - Percent Row: Calculates a percent based on the values.
    - Percent Column: Calculates a percent based on the values.
    - Total Column: Totals all of the values.
    - Total Row: Adds a new row containing a total of the values.

When you choose one or more of the methods above, an abbreviation for that method will be prepended to the column header, unless Sum, First, or Last is the only method selected.

5. The character specified in the Separator field is automatically added between concatenated strings. This option is only available when concatenating strings.
6. The Field Size field is the maximum field length (in characters) for concatenating strings. If the string is larger than the size specified, a warning appears in the Results window and the data is truncated. Visit the Data Types article for information on string data types

N·E·A·T
प्रौद्योगिकी के लिए राष्ट्रीय शैक्षणिक सहयोग
National Educational Alliance for Technology

अखिल भारतीय तकनीकी शिक्षा परिषद्
All India Council for Technical Education

EduSkills®
Nation Building Through Skills

# Certificate of Virtual Internship

This is to certify that

## Badarinadh Bonam

Vishnu Institute of Technology

has successfully completed 10 weeks

**Data Analytic Process Automation**

during May - July 2023

Supported By alteryx SPARKED

**Olivia Duane-Adams**
Chief Advocacy Officer (CAO)
Alteryx

**Shri Buddha Chandrasekhar**
Chief Coordinating Officer (CCO)
NEAT Cell, AICTE

**Dr. Satya Ranjan Biswal**
Chief Technology Officer (CTO)
EduSkills

Certificate ID :48cfee5f7752fef1c2c3ca57deac6f42
Student ID :STU643ad68b460df1681577611