# Tour:

* Tool Palette
  + In/Out
  + Parse
  + Preparation
  + Join
  + Transform
  + Reporting
  + Etc.
* Canvas
  + Link Tools by Output anchor to input anchor
  + Way to have a took in Canvas
    - Drag and drop from Tool Palette
    - Search and drag
    - By Right Click on Insert and select the tool
* Configuration Window
  + Tools show red alert when the configuration of the tool is incorrect
  + Show preview of 100 records
* Results Window
  + Populate below information when running workflow
    - Links to Input
    - Summarize the changes applied
    - Time stamp of workflow start and complete
    - Error messages and warnings
  + After running the workflow, we can view the results of each component in results window
* Saving Workflow
  + Workflow will be saved in .yxmd
  + Export the workflow by click on Options-> Export workflow; this will create .yxgp, which is Alteryx package. Which contains .yxmd and any assets (input data sets, macros)

Tools Information

|  |  |  |  |
| --- | --- | --- | --- |
| Tool Name | Action | Steps | Comment |
| Select Tool | Change Data Type |  |  |
| Join | Bend/Join the Data | 1. Drag and Drop to Canvas 2. Configure Left and Right fields mappings 3. We can use configuration window and change below (Embedded Select) 4. Select Required Fields 5. Change Data Type 6. Data Type Size 7. Rename Columns 8. Results 9. Right – Having in Right but does not have in Left 10. Left - Having in Left and does not have in Right 11. Join – matched records in left and right | 1. Blend data together using Key Fields 2. This tool has two inputs one is Left and the other is Right 3. This tool has 3 Outputs. Right, Left and Join |
| Summarize | Aggregation | 1. Configuration Window 2. In the field select required column 3. In the Action choose the required aggregation |  |
| Filter | Filtering - Test column value | 1. Filter column value by condition and divide into two streams 2. Two Outputs 3. T - True Condition 4. F - False Condition 5. Filters 6. Basic 7. Custom Filter |  |
| Sort | Sort the data | 1. Configuration Window   Provide column name and the order type |  |
| Output Data | Write output stream | 1. Configuration Window   A)Write to File or Database |  |

## Data Types:

* Columns are categorized into 5 types
* Most of the times we can use ‘ auto field tool’ too automatically adjust the data types by alteryx designer
* Sometime when we need to decided our own data type then we can use Select Tool
  + Strings
    - Qualitative numbers should be String data type
    - These cannot be measured
    - Subtypes
      * Variable
      * Fixed
    - String & Wstring are set in length
    - V\_String are variable length
    - Wstring can capture international characters
  + Numeric
    - Quantitative numbers
    - 7 Subtypes
      * Whole Numbers
        + Byte - 0 to255
        + Int16 - -(2 p 15) to (2 p 15)-1
        + Int32 - -(2 p 31) to (2 to 31)-1
        + Int64 - -(2 p 63) to (2 p 63)-1
      * Decimals
        + Fixed
        + Float - +/-(3.4E+/-38)
        + Double - +/-(1.7E+/-308)
  + Date & Time
    - DateTime tool convert string to data and date to string
  + Boolean
    - Two logical values
    - 0 - False
    - 1 - True (non Zero)
  + Spatial (Geo)
    - Point represented by X and Y coordinate
    - A Line or Polygon
* We can see meta data of any tool by click on the tool and in the results window click on metadata button
* It has
  + Name
  + Type
  + Size
  + Source
  + Description
* Based on the input and output tool selected designer will show available/supported data types

## Inputting Data:

* Input Data Tool
  + Connect to verity of database and file types
  + It can be found in In/Out Tool Palette
  + It can be found also in favorite tool category
  + We cannot use Manage data connection switch to user dynamically other DB’s via aka update
  + Dynamic Input
    - All files in the same directory and the same columns and data types
      * We can just use whildcard in the connection string
      * It works even if the files are in sub directories, then in that case we have to check the ‘Search SubDirs’ option
    - It also supports zip formats
  + File Types
    - CSV
      * Record Limit
        + Limit the number of records to show
      * Output file Name as field
        + This is to show the file name in the results window while running workflow

No

File Name Only

Full Path

* + - * Delimiters
      * First row contain field names
      * Field Length
      * Start data import on line
        + Avoid importing extraneous information from the data source
      * Code Page
        + Some data encoded with particular format i.e. UTF 8
    - Excel Workbook
      * Import Options
        + Select Sheet
        + Select Named Range
        + Import only the list of sheet names
      * Options
        + Record Limit

Records to show in designer

* + - * + Table or Query

Selected sheet name will show here, we can also change sheet name here

* + - * + Output file Name as field

This is to show the file name in the results window while running workflow

No

File Name Only

Full Path

* + - * + First Row contains data
        + Start the data import line
    - YXDB (Alteryx database file)
      * Most efficient file type reading data into Alteryx
      * It is compressed, storage value in same name column types, its quick
      * Options
        + Row Limit
        + Search SubDirs
        + Output file name as field
  + Preview of the tool shows 100 records, after running the flow it shows complete records in the results window

## Manually Entering Data:

* Text Input Tool
  + It is used to enter data manually
  + Configuration Window
    - It provides blank table to provide customization
    - It can also read existing files
    - We can copy past also
    - Data lives with in the workflow
    - Data values are static
    - Can be shared by sharing the workflow
    - Data automatically categorized into smallest type and size
    - Blank data in the text input tool will replace with null value
    - Alteryx will throw errors in case of any file format issues, but we can still read the file by check on the ‘Treat Read Errors as Warnings’
    - When we give delimiter as “\0” then Alteryx consider there is no delimiter and read entire line to one column

## Formatting Data:

* Select Tool
  + Allow changes to column meta data
    - Name
    - Data Type
    - Size
    - Description
  + We have the ability to save the current select configuration to a file .yxft and that can be loaded later user.
  + When number are showing in the thousands with comma, then to represent it as number we can select ‘use comma as decimal seperator’
* Join Tool
  + Contains embedded select

## Viewing Data:

* Browse Tool
  + Too see larger data sets (CTRL + SHIFT +B)
  + Needs temp space and physical memory to visualize your data
  + # Records, # Fields, # Bytes
  + Recommendation is to add the browse tool after input to observed the quality of the data
  + If you don’t select the Browse Tool when you are working in the workflow it will not use any memory until you select the Browse Tool to render the reports
  + Using Data Profiling before browse tool help solve issues with rendering huse data
  + Fewer than expected records?
    - Join is case sensitive, if you have different cases in two different datasets then it is not going to join
    - If you have leading or trailing spaces then also it will not join the data
  + Data Quality
    - Not OK
    - NULL
    - EMPTY
    - OK
  + Pros
    - In-depth view of data
    - Can view entire data set
  + Cons
    - Memory Intensive
    - Increase time and processing requirements
  + We can use input and output anchors to view the data quickly
  + Input anchor
    - Prior to apply changes
  + Output anchor
    - After applying the changes

## Blending data with Unions:

* Union Tool
  + Combine data sets
  + First input connected to the tool determine the output column names and datatypes
  + It shows the #numbers with the order of connection
  + We can double click on the #number to rename
  + Combining reporting elements vertically
  + Options
    - Column Names
      * Aligning column data based on names vertically is default behavior
    - Column Position
      * It combines based on the order of the columns 1st, 2nd etc..
      * In this case the columns should be in the same order
    - Manfully
      * We can shift the columns left/right manually
  + When number of columns are different in each source we can choose below options
    - Warning - Continue processing records
    - Error - Stop Processing the records
    - Ignore – Continue processing the records
  + We can choose output type as well
    - Output all fields
    - Output common subset of fields, it is by default seleted
  + We can select which data to show on order by using select option in configuration window
  + Union tool does not support combining based on data type
  + Does union tool eliminate duplicates?

## Filtering Data:

* Filter Tool
  + It supports two types of configuration
  + It is part of preparation category
  + Remove Rows with Single Null Value
  + This tool support connectors from Date tool to use and filter the data
    - Basic
      * Simple queries
      * Evaluate single condition in one column of data
      * Three parts
        + Column name
        + Operator

Will display the list accordingly to the column name selected

* + - * + Condition to test
    - Custom
      * Complex queries
      * More than one condition across multiple columns
      * Selected basic options replicate in the custom box
      * We can give multiple statements by AND / OR

## Joining the data:

* Join Tool
  + It has two anchors the left and right
  + Type of joining
    - Join by Record Position
      * First Row in the left input horizontally align data with right input
      * It produces quick results
      * But be careful when the columns have wrong sequence of data will cause for wrong output
    - Join by Specific fields
      * Both data inputs should have common data column
      * Values will horizontally align
      * More than one column can be selected for more restrictive match
  + Provides embedded select window, it provides features of select tool
    - Remove Columns
    - Rename Columns
    - Reorder the Columns
  + Only same data type filed can join together
  + Output anchors
    - J
      * Linked data from both inputs
    - L
      * Data from the left input that do not join with the any data in right input
    - R
      * Data from the right input that do not join with the any data in left input

## Soring the data:

* Typically we sort 3 types of data strings, numbers and dates
* These can be sorted in two ways ascending order or descending order
* When in case of applying multi row functions, the order of row is very important sort tool help us solve that.
* Listing records for chronological order for the reporting
* When using summarizing tool, it helps to pick first or last based on sort order
* Use of Dictionary Order - Assume you have filed with values but the data type of the filed is String, when we sort it sort by all 1’s fist irrespective of whether it is 1 or 10 or 100. When we use dictionary Order, it solve this purpose of sorting by applying its numeric values
* Sort Tool
  + When duplicate values found it preserve the order in the data set
  + Multiple columns can be selected in the sort option
  + By default string values will be sorted based on the ASCII value of the character
  + String contains punctuation (single quotes) followed by Upper case and then lower case will appear at the top of the data set 🡪 for Ascending order
  + When enable the order by ‘dictionary’ it shows based on dictionary and ignore case

## Writing the data:

* Output data tool
  + It does only have single anchor
  + Output 3 types
    - CSV
      * Select file type as ‘Comma delimited text files’
      * All data types will store as string data type
      * Unsupported columns will be dropped from the output like spatial objects
      * Options
        + Max Records per file

Limit the number of rows exported in a single file

When data container more rows, multiple files will be created

* + - * + Delimiter
        + Code Page
    - Excel Workbook
      * File type should be Microsoft excel
      * Select File from the configuration window
      * Provide sheet name
      * Spatial objects are not supported
      * Options
        + Max Records per file

Limit the number of rows exported in a single file

When data container more rows, multiple files will be created

* + - * + Output options

Create new sheet

Append to existing sheet

Overwrite sheet (Drop)

Overwrite file (remove)

* + - * + When creating first time we needs to select ‘Create new sheet’ in the subsequent runs we may need to modify
    - Alteryx database
      * No limit on number of rows it store
      * Options
        + Max Records per file

Limit the number of rows exported in a single file

When data container more rows, multiple files will be created

* + - * + Save Source & Description

By default it stores source and description of the source in metadata

This useful to understand the oration of data

* + Run workflow(CTRL + R)
  + Output to multiple files by options
    - Take File/Table name from field
      * Options
        + Append suffix to file/table name
        + Prefix to file / table name
        + Change file /table name
        + Change entire path
      * Field containing file name or part of file name

## Excel:

### Connecting to Multiple Sheets at once:

* Dynamic Input tool
  + Facilitate input multiple sheets
  + It does have input and output anchor
  + It needs names of multiple sheets as input to read in
  + Standard input tool can generate list of sheet name and file path
  + Reading multiple files are supported only in .XLSX file
  + Template
    - Number of columns in each sheet should be match, in case of no match it will not import the data
    - Columns can be in any order when the names are in match
    - If column names are different data will be read positional
    - First selected sheet act as an template
  + In developer menu you can find dynamic input
* Lookup
  + Vlookup or Replace function
  + Find Replace Tool
    - Identify column in main data set to user
    - Identify look up table
    - Find Anchor
      * Data file we want to search
    - Replace Anchor
      * Lookup table contain replacement value
    - Data set connected to the find anchor will be the output
    - Options
      * Find
        + Search

Beginning of the column

Search for the very first character(s) in the string

Any part of the column

Entire Column

* + - * + By default it is case sensitive, if we do not want this Case insensitive box should be selected

|  |  |
| --- | --- |
| Find Replace Tool | Join Tool |
| Flexibility | Requires Exact Match (Case, spelling, punctuation, white space) |
| Allow Case insensitive |  |
| Placement with in cell |  |
| Direct Replacement of Data | Append Columns only |
| String Data Types Only | Accepts multiple data types |
| No Embedded Select | Embedded Select |
| Returns Single Match | Return all identified matches |

### Separating Data into Columns and Rows:

* Text to Columns Tool
  + It can be found in the parse tool pallet
  + Multiple delimiters can be specified
  + Options
    - Column to Split
    - Delimiters
    - Split to columns
      * Number of Columns
      * Extra Characters
        + Leave extra in the last column (Default)
        + Drop Extra with Warning
        + Drop Extra without warning
        + Error
      * Output Root name
    - Advanced Options
      * Ignore delimiter in quotes
      * Ignore delimiter in single quotes
      * Ignore delimiter in parenthesis
      * Ignore delimiter in brackets
      * Skip empty columns
    - Split to Row

### Changing Data Layouts

* Transpose
  + Pivots horizontal data to vertical
  + At least output two columns of data
  + Column called name contains headers from incoming data
  + Output records = number of headers X number of rows (4 X 3 = 12)
  + We can select what data can be transposed from the ‘Data Columns’ Option
  + Key columns will be excluded from the transpose process
  + Key column maintain current orientation
  + Columns shows ascending based on alphabetically

Input Data:

|  |  |  |  |
| --- | --- | --- | --- |
| **APPLES** | **SPUDS** | **CARROTS** | **TOMOTAS** |
| Green | Small | Yellow | Roma |
| Red | Medium | Purple | Cherry |
|  | Large | Orange | Heirloom |

Output data:

|  |  |
| --- | --- |
| **NAMES** | **VALUE** |
| **APPLES** | Green |
| **SPUDS** | Small |
| **CARROTS** | Yellow |
| **TOMOTAS** | Roma |
| **APPLES** | Red |
| **SPUDS** | Medium |
| **CARROTS** | Purple |
| **TOMOTAS** | Cherry |
| **APPLES** |  |
| **SPUDS** | Large |
| **CARROTS** | Orange |
| **TOMOTAS** | Heirloom |

Key Column (SPUDS):

|  |  |  |
| --- | --- | --- |
| SPUDS | NAMES | VALUE |
| Small | **APPLES** | Green |
|  | **CARROTS** | Yellow |
|  | **TOMOTAS** | Roma |
| Medium | **APPLES** | Red |
|  | **CARROTS** | Purple |
|  | **TOMOTAS** | Cherry |
| Large | **APPLES** |  |
|  | **CARROTS** | Orange |
|  | **TOMOTAS** | Heirloom |

Key Column (SPUDS, CARROTS):

|  |  |  |  |
| --- | --- | --- | --- |
| **CARROTS** | **SPUDS** | **NAME** | **VALUE** |
| Yellow | Small | **APPLES** | Green |
| Yellow | Small | **TOMOTAS** | Roma |
| Purple | Medium | **APPLES** | Red |
| Purple | Medium | **TOMOTAS** | Cherry |
| Orange | Large | **APPLES** |  |
| Orange | Large | **TOMOTAS** | Heirloom |

* Cross Tab
  + Vertical data to horizontal
  + At least two columns must be present in incoming data set
  + Cross tab tool required pattern in order to pivot
  + We can create a conditional column that will serve as pattern
  + At least two columns must be present in the incoming data set
    - One column server as data headers
    - Other will serve as values
  + Only one column can be specified for headers and values
  + It removes duplicate values and make them aggregate
  + Data would be aggregated based on column type
  + Usually it will generate only one row of data , but when we need multiple rows, we can use group column
  + Any values which are not selected in the cross tab configuration will be dropped
  + String Type
    - Concordant or user first, last value
  + Numeric Type
    - Sum, Avg, Fist and Last

Input Data:

|  |  |  |
| --- | --- | --- |
| **FRUIT** | **TYPE** | **QUANTITY** |
| Apple | Green | 6 |
| Apple | Red | 9 |
| Spuds | Small | 12 |
| Spuds | Medium | 15 |
| Spuds | Large | 8 |
| Carrots | Yellow | 3 |
| Carrots | Purple | 5 |
| Carrots | Orange | 18 |
| Tomatoes | Roma | 10 |
| Tomatoes | Cherry | 8 |
| Tomatoes | Heirloom | 5 |

Output Data: (Column change header =FRUIT, Values for new columns = QUANTITY, no selection from group by data, aggregation as sum):

|  |  |  |  |
| --- | --- | --- | --- |
| APPLE | CARROTS | SPUDS | TAMATOES |
| 15 | 26 | 35 | 23 |

* Things to remember
  + Incoming data set have multiple data types, values would convert to V\_String
  + Special characters and spaces in header with underscore
  + After cross tab we may needs to rename the columns
  + Cross tab groups headers in ascending order

## Sampling the data:

* Sample Tool:
  + There are 6 methods to apply samples
  + It drop the records which does not needed
  + We can also apply group column, when we select group column n rows will be sampled for each group item in the selected column
  + It removes in the sampling process but not in the output
  + If we want to retain all then filter tool is appropriate
  + We can do flexible parsing
  + Stripping (often Excel) formatting from input tables
  + Quick appending min/max or chronological first/last in date types
* Record ID
  + Maintain the record number of the row
  + It always remember the original record it, it will not update with the transformations applied in any steps
  + As it preserve the original position, it takes advantage there

## Removing Duplicate Values:

* Unique Tool
  + Unique tool has two output anchors
  + U - Unique values only
  + D - Output values determined to be duplicate
  + It process data from top down
  + We should use sort tool before unique to ensure right data
  + Unique tool is case sensitive
  + In order to resolve the case sensitivity and spaces in the data, we can use data cleansing tool before unique tool to solve issue.
  + first record in each unique group is sent to the U output stream while the remaining records are sent to the Duplicate output stream
  + Using before a join to ensure you don’t get duplicate records
  + Unique tool provide sorted output based on the field we selected for unique

## Summarizing data:

* Summarize Tool
  + Can be used for Grouping
  + Summarize tool is case sensitive
  + Only columns in the Actions are applied and kept in output
  + It also can be applied to spatial data type

## Writing Expressions:

* 13 tools can use expressions
* Formula Tool
  + Used to create new column of data
    - We have to name the new column
    - Default data type of V\_WString is assigned
  + Or modify existing column
    - Data type and size cannot be changed in formula configuration.
  + Expressions Editor
    - Can be entered Variables, Functions and syntax
    - X - to get existing columns or typing [ bracket
    - Data preview show first row from the data input dataset
  + It applies the formula to all the rows
  + We can create multiple new columns in the formula tool
  + Expressions in the Formula tool executed sequentially
  + CTRL + SPACE to see full list of functions
    - Functions in Black Color
    - Columns in Blue Color
    - Constraints in Green Color
  + To see the data preview at least one time the workflow has to run
  + CTRL + to increase the size of the formula

### Writing Conditional Statements:

* Basic conditional statements
  + IF - THEN - ELSEIF-THEN-ELSE - ENDIF
  + Inline IF statement
    - IIF (Condition, True value, False Value)
  + SWITCH(Value, Default,Case1,Result1,Case2,Restul2,CaseN,ResultN)

### Using NULL and Empty in expressions:

* Null and empty represent two distinct ways of showing no data is present
* Null can be found in specialized
* Isnull() is used to evaluate null values in the column
* Isnull() does not evaluate blank values
* Isempty() will test both null and empty values

### String Functions:

* Remove white spaces
  + TrimLeft()
  + TrimRight()
  + Trim()
* Replaces
  + Replace() – Replace all occurrences
  + ReplaceFirst() – Replace only first occurrence
  + ReplaceChar() – Replaces single character, when multiple characters are provided then for each character replacement will be done
* Standardize String Values
  + Upper()
  + Lower()
  + TitleCase()
* Function Types
  + Conditional Statements
  + Null & Empty Values
  + String Values
  + Numeric Values
  + DateTime Values

### Numeric Functions:

* CEIL(x) - round up to the nearest integer
* FLOOR(x) - Round Down to the nearest integer
* SmartRound(x) – Rounds to the nearest multiple dynamically decided by the Designer
* Round(x,mult) – Round to the nearest multiple specified

### Date & Time Functions:

* Date
  + Standard format would be YYYY-MM-DD
* Time
  + HH:MM:SS
* Date Time
  + YYYY-MM-DD HH:MM:SS
* DateTimeAdd(dt,interval,unit)
* DateTimeFormat(Date,Format)
* DateTImeDiff(date,date,unit)
* Important functions
  + DateTimeParse
    - takes a String data type that represents a date and/or time and returns the value in a DateTime data type
  + DateTimeFormat
    - takes a Date, Time, or DateTime data type and returns a string in the format designated

### Spatial Functions:

* It does user longitutue and latitude and convert them into points
* Once they are convert to points we can draw a line from start to end using line formula in spacialobj
* Most probable questions it answers is how special objects related to each other

### Using Regular Expressions:

* These can be found in String function library
* REGEX\_Replace(String,pattern,replace)
* Marked groups
  + These are the data identified by the regular expressions
  + We can keep the identified text with in column
* REGEX\_Match(String, pattern)
* REGEX\_CountMatches(String ,Pattern)
* Output from regexmatch is either 0 or -1
* Case sensitivity can be forced by optional parameter

### Multi Row Formula:

* Similar to formula tool, we can create new column or replace existing column
* Both have expression editor
* Both process data from top down
* Expressions are applied to row in order
* Multi row can use values from other rows
* Can only apply one expression per tool
* Fill Cells
  + Duplicate
  + Serial Fill
* When apply group by the sequence will re-start at each new value in group
* When previous is not found value of 0 will be used, it is default, we can change this

### Multi Field Formula:

* Supports only one expression per tool
* It will create new columns by default, we can uncheck the check box provided to overwrite this behavior
* We can apply the formula by data type, we do not have option to select multiple columns of different data type
* Allow to change the data type of existing column

# Input data:

* Input data tool
* Download Tool
  + Connect to HTTP/FTP/SFTP
  + Web Scrap
  + REST API’s

### Select Dates:

* Use date tool to capture the user input
* To test the values, select empty area in canvas
* Go to workflow tab in the configuration window
* You can see ***Question*** in the Type that is representative of date tool.
* You can enter the required date in the Value in text format
* You can apply in the formulas like Question.<Name>
* There is an another option is to enter values
  + Select the Date tool
  + Click on wheel button in the configuration window
  + In the text box enter the required values
* You can drag and connect of few supported tools like Filter, when you drag and drop it shows the number with #1, that can be used directly in the formula window

# File Types:

* Yxzp - Workflow Packages, which bundles the source files and output files together with the workflow
* Yxmd - Workflow, it does not copy any source/output files
* Yxmc - Macros
* Yzwz – Apps
* Yxft - Field Type Files (Field Configuration)

# Detour Tool:

* Divert the path of the workflow either to right side or left side
* By default Right side is selected
* Detour tool should finally connected to either Output tool or Detour End tool
* All data flows through either R or L
* Detour end tool brings the back to single stream

## Dynamic Rename:

* Rename field names
  + Use first row as header
  + It has two anchors
  + We can use anchors to supply data into Left and Field names into right
  + We can add Prefix or Suffix to the filed names
  + There is an option to user the formula, but it only applies to the headers not to the data of the column

## Dynamic Rename:

* Append the column

## Dynamic Select:

* One common use case to remove blank column by Select Based on the datatype of the column
* Multiple ways it applies to columns

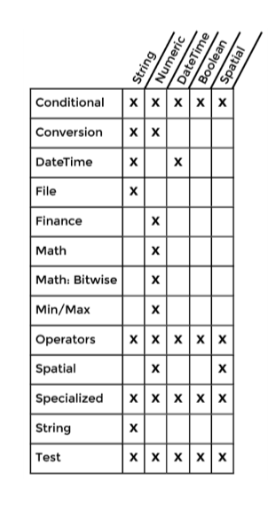
## Tile Tool:

* It provides tiles based on the groups
* It provider sequence numbers based on groups configured

## Generate Rows Tool:

* Used to generate rows based on condition
* Common case is to generate dates between two dates
* It have Initialization, Condition (Termination) and increment conditions

## Functions Application:



## Summarize:

* Grouping, Summing, Counting, Spatial processing, string concatenation etc,.
* It supports every data type – numerics, strings, spatial objects, behavior profiles, and even reporting elements

## Output Tool:

* Can write multiple files of any format
* It can also directly output to Tableau Server and PowerBI
* Output to your Local and Network Drives
* Output to your Database Connections