

# **DDP ASG 2**

***Bus arrival timings with Alteryx***

**Clayton Thia**

# **Presentation content**

**01. Workflow**

**02. Findings**

**03. Demo / QnA**

# Workflow

01

Data preparation

02

API Call

03

Further data  
preparation

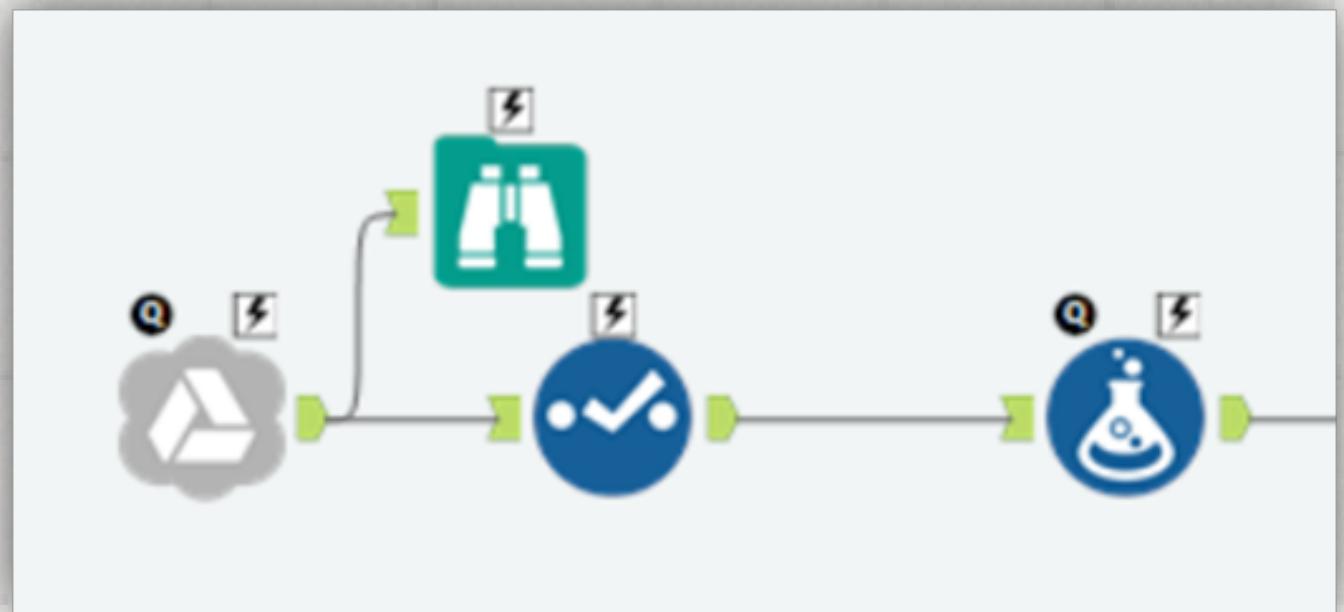
04

Data display

# Data preparation

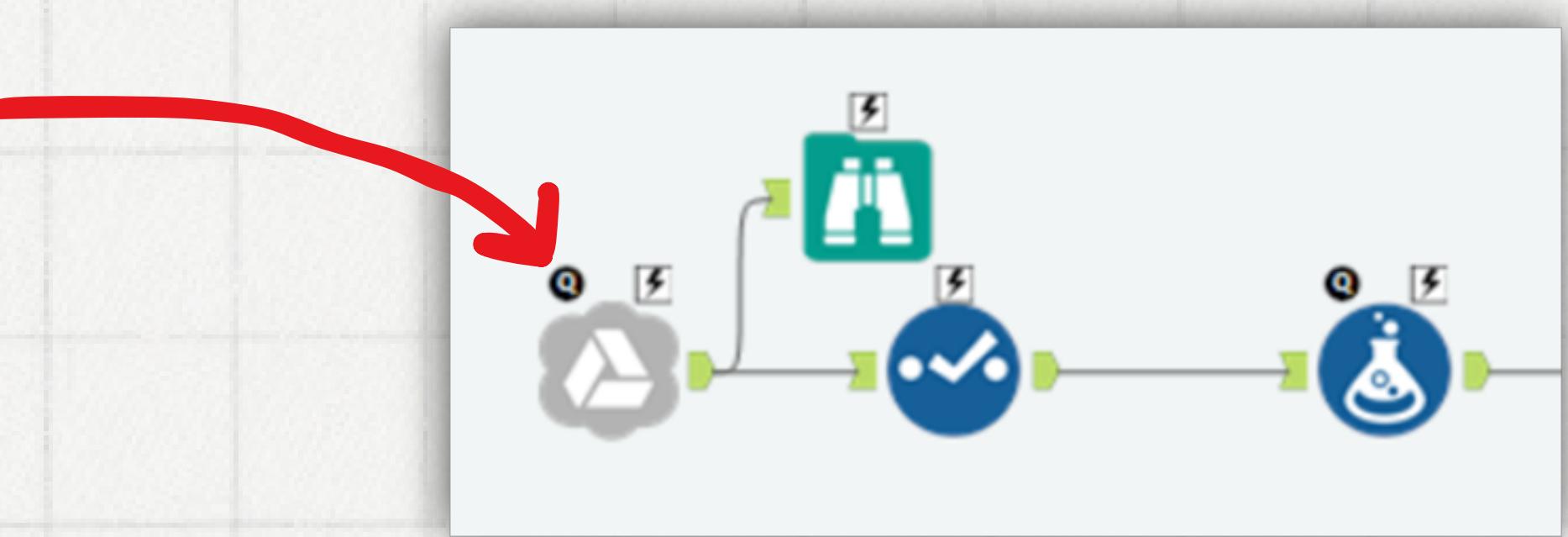
# Data preparation

- Prep model to provide proper API call based on bus stop code input



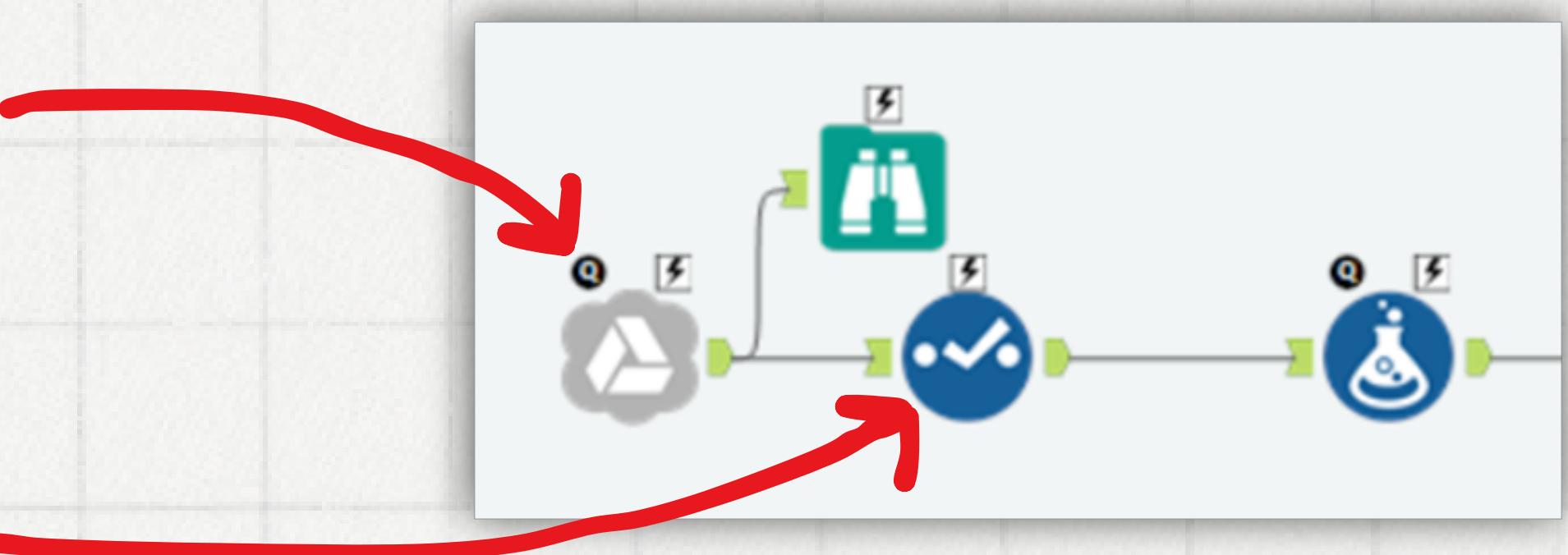
# Data preparation

- Prep model to provide proper API call based on bus stop code input
- Model gets data from Google Sheets
  - Incomplete URL



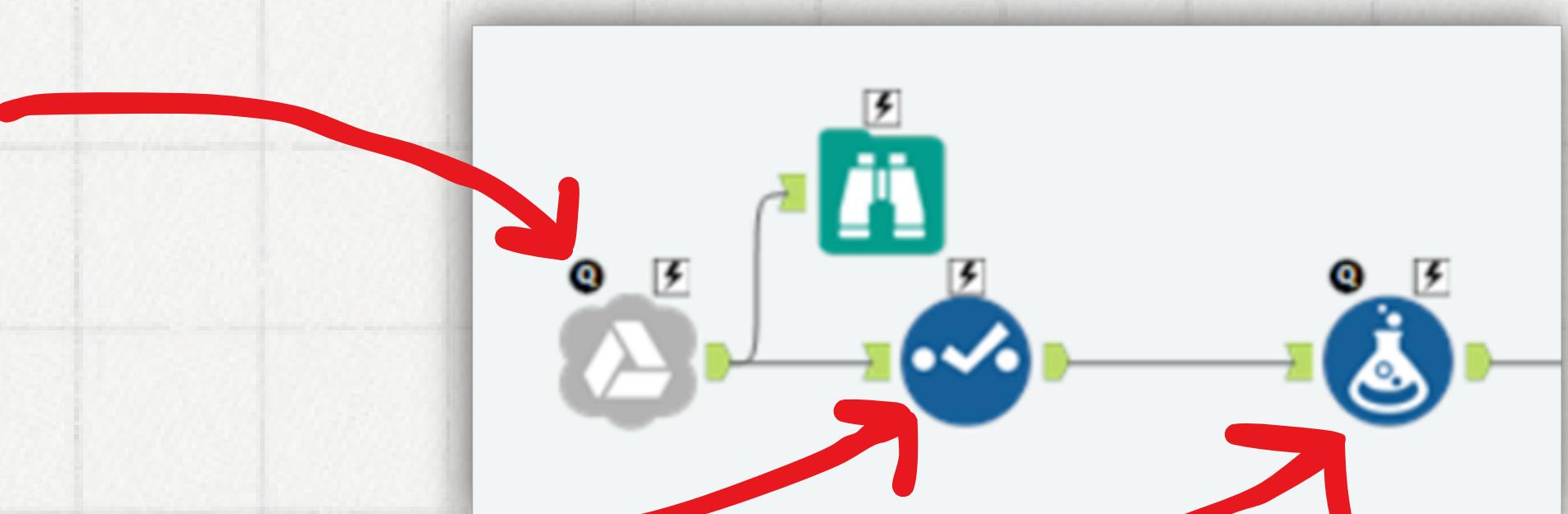
# Data preparation

- Prep model to provide proper API call based on bus stop code input
- Model gets data from Google Sheets
  - Incomplete URL
- Data type for bus stop code column changed to string



# Data preparation

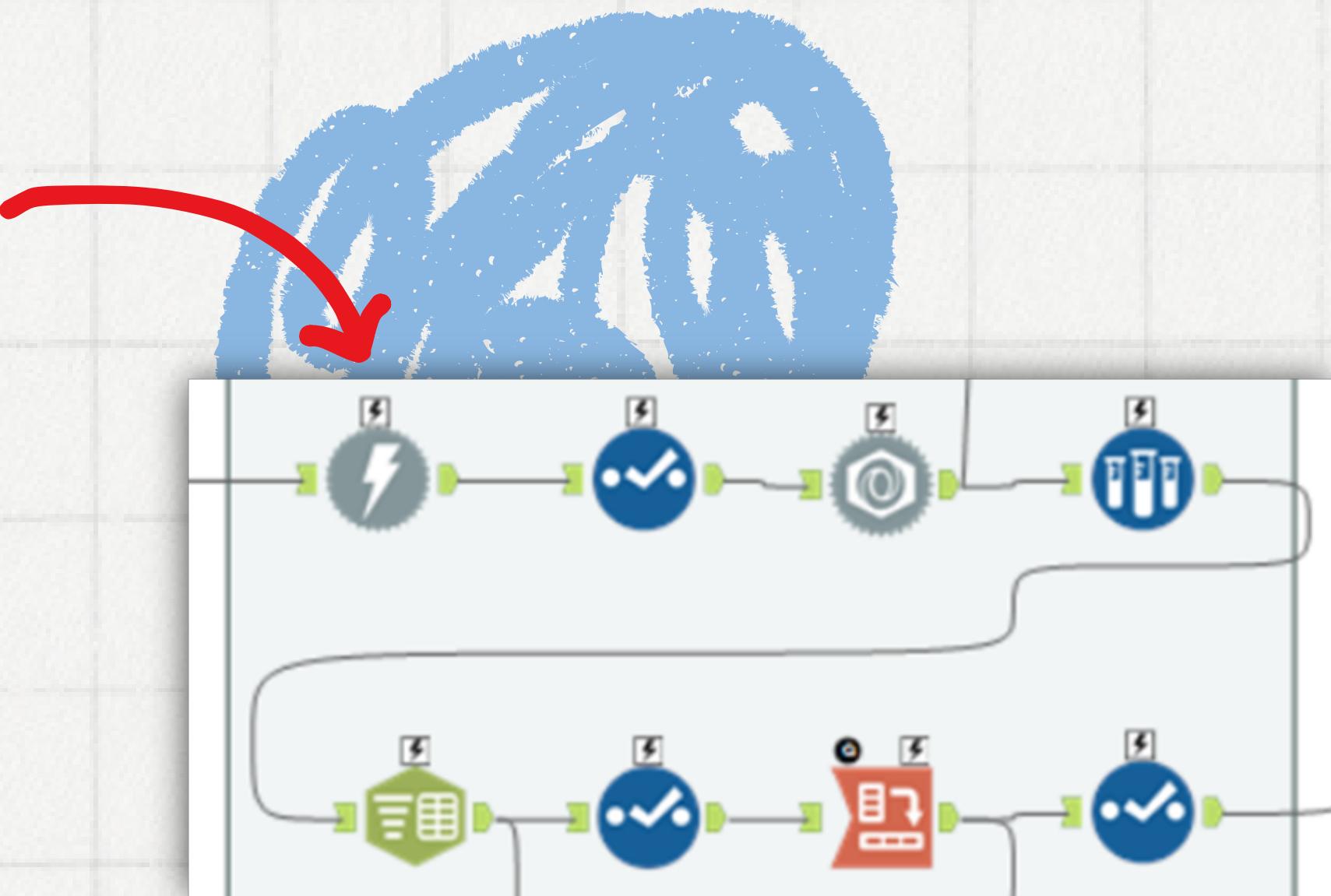
- Prep model to provide proper API call based on bus stop code input
- Model gets data from Google Sheets
  - Incomplete URL
- Data type for bus stop code column changed to string
- Completed URL by appending bus stop code to incomplete URL



# API Call

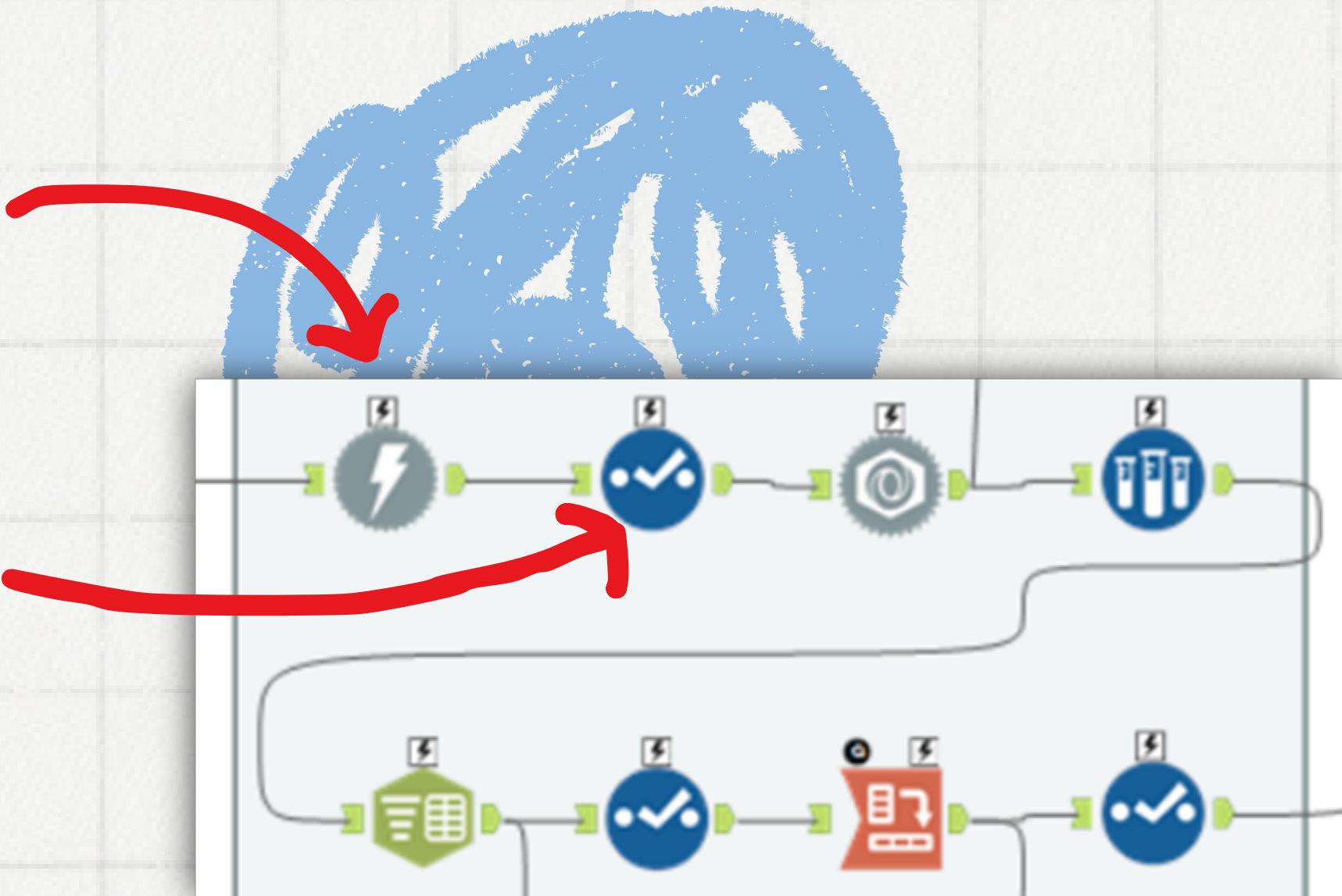
# API Call

- Download output from API call and transform data into proper structure



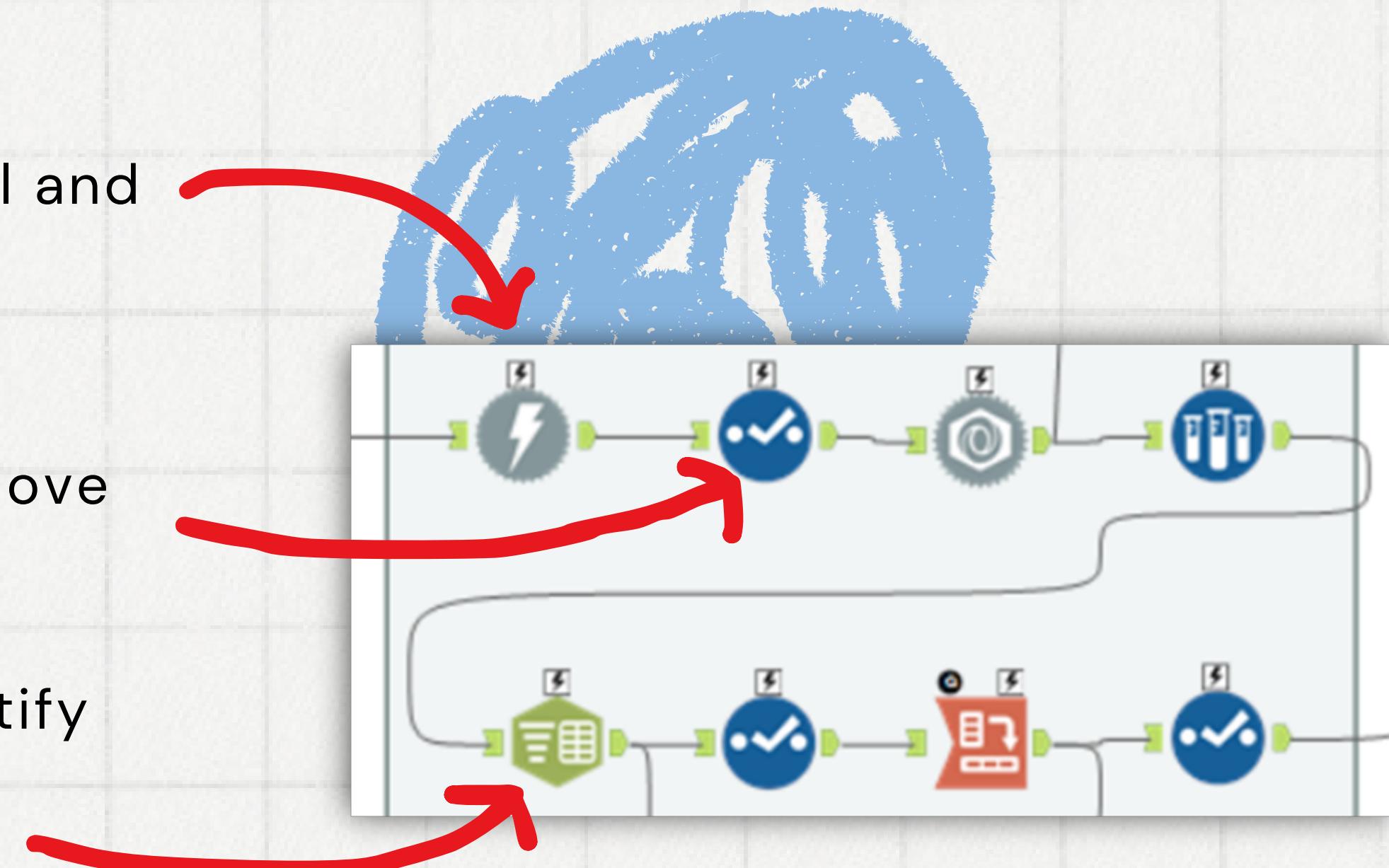
# API Call

- Download output from API call and transform data into proper structure
- Select tools were used to remove unnecessary columns



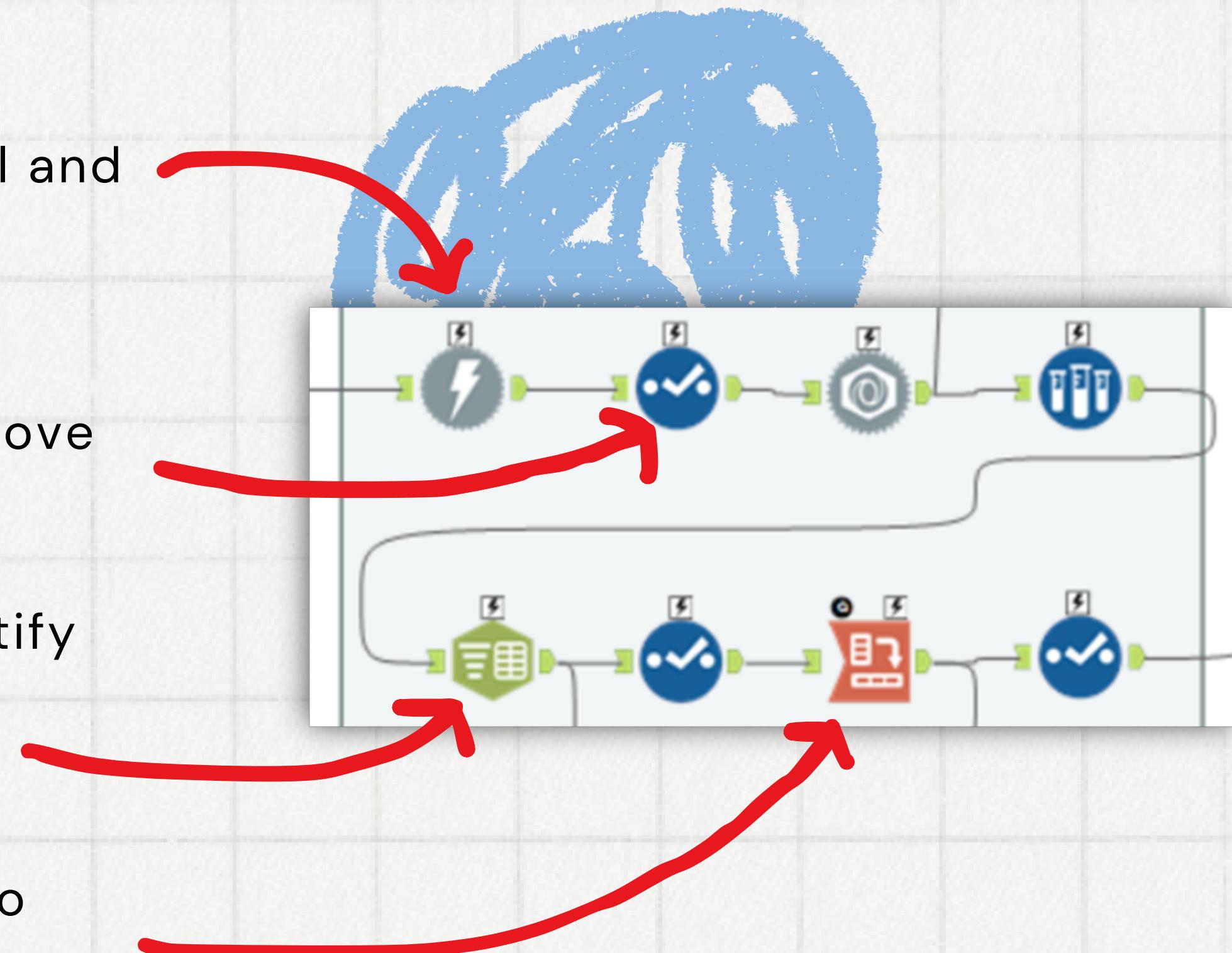
# API Call

- Download output from API call and transform data into proper structure
- Select tools were used to remove unnecessary columns
- Text to Columns used to identify features for different service numbers



# API Call

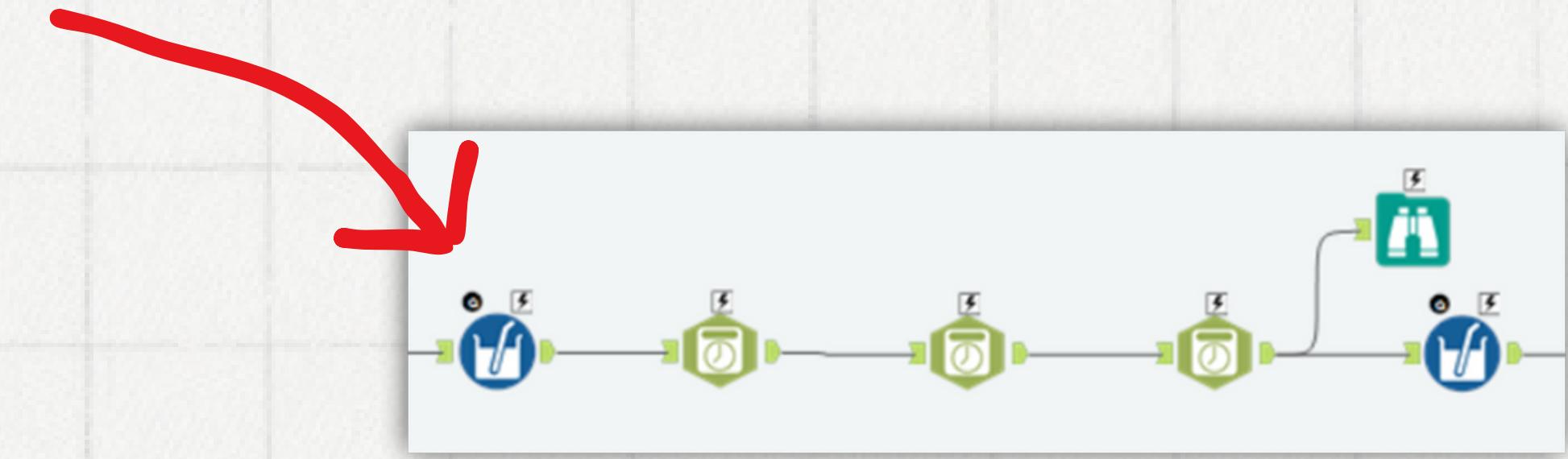
- Download output from API call and transform data into proper structure
- Select tools were used to remove unnecessary columns
- Text to Columns used to identify features for different service numbers
- Cross Tab to convert rows into columns



# **Further data preparation**

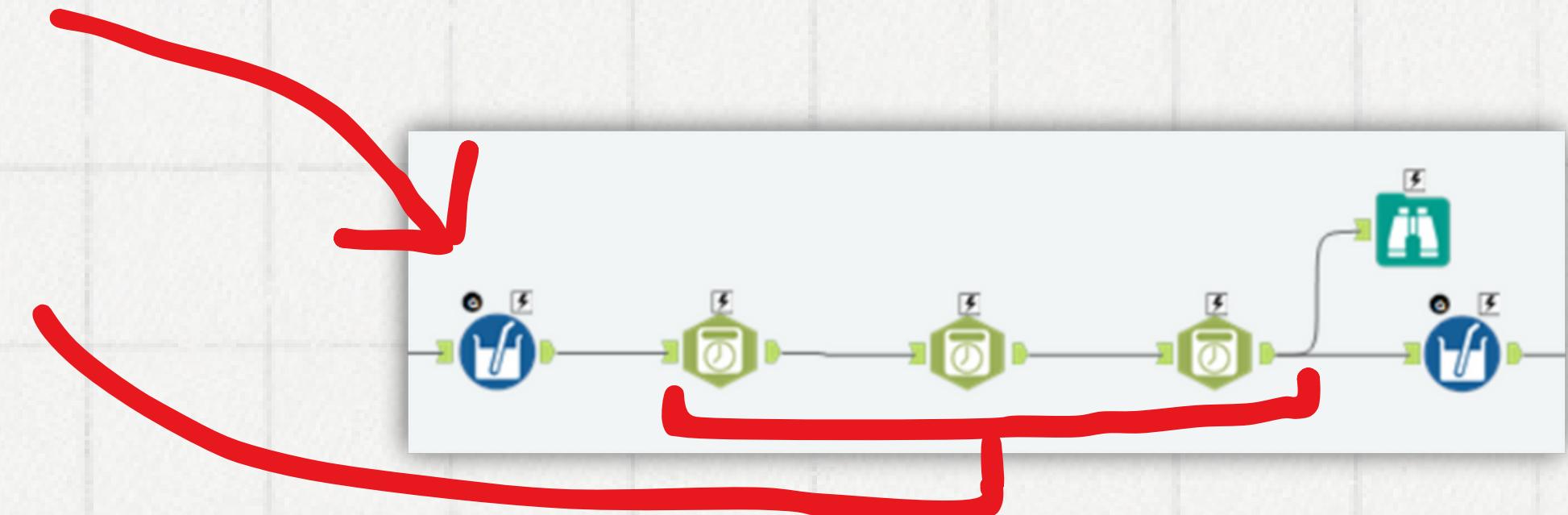
# Further data preparation

- Remove time-zone offset from time text



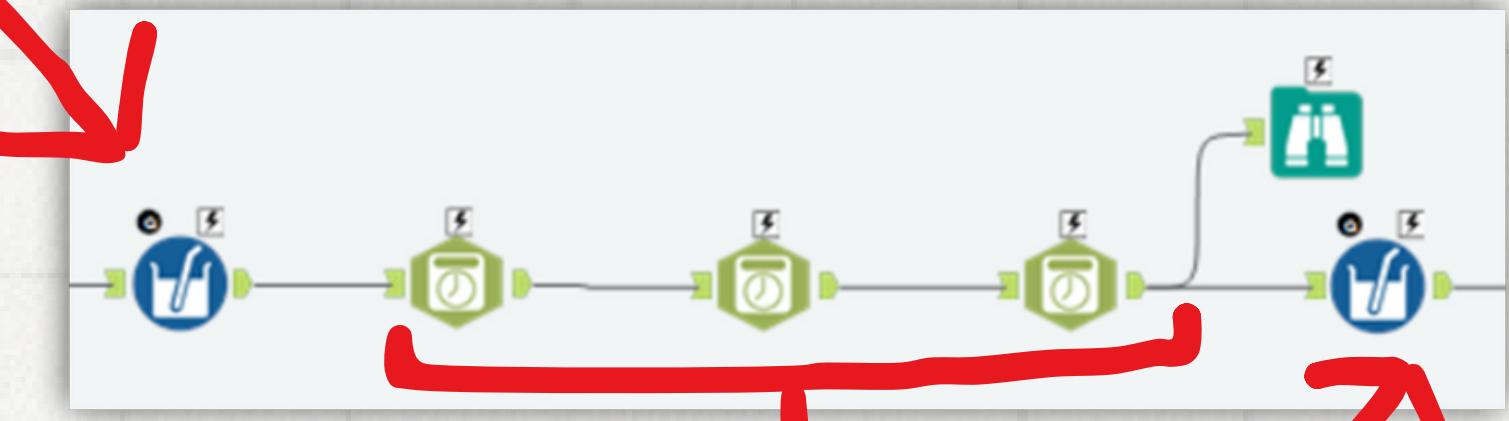
# Further data preparation

- Remove time-zone offset from time text
- Convert columns to DateTime



# Further data preparation

- Remove time-zone offset from time text
- Convert columns to DateTime
- Extracting DateTime information
  - Minutes to arrival



# Data display

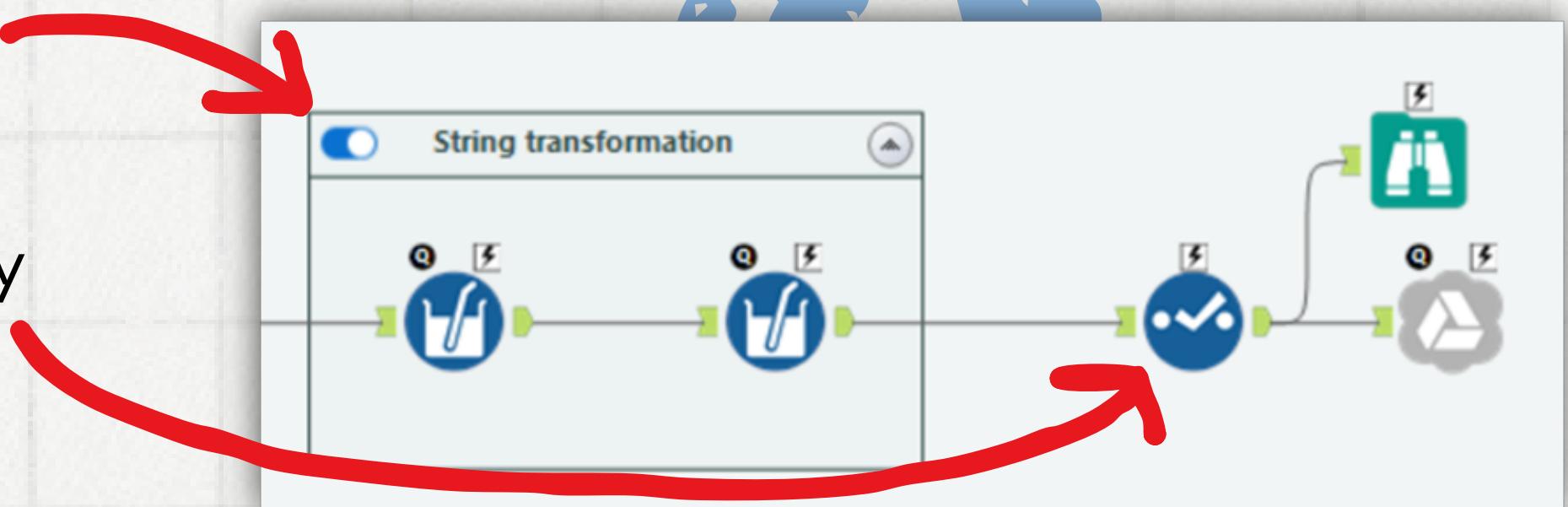
# Data display

- Replacing abbreviations
  - E.g. SD > Single (Single deck)



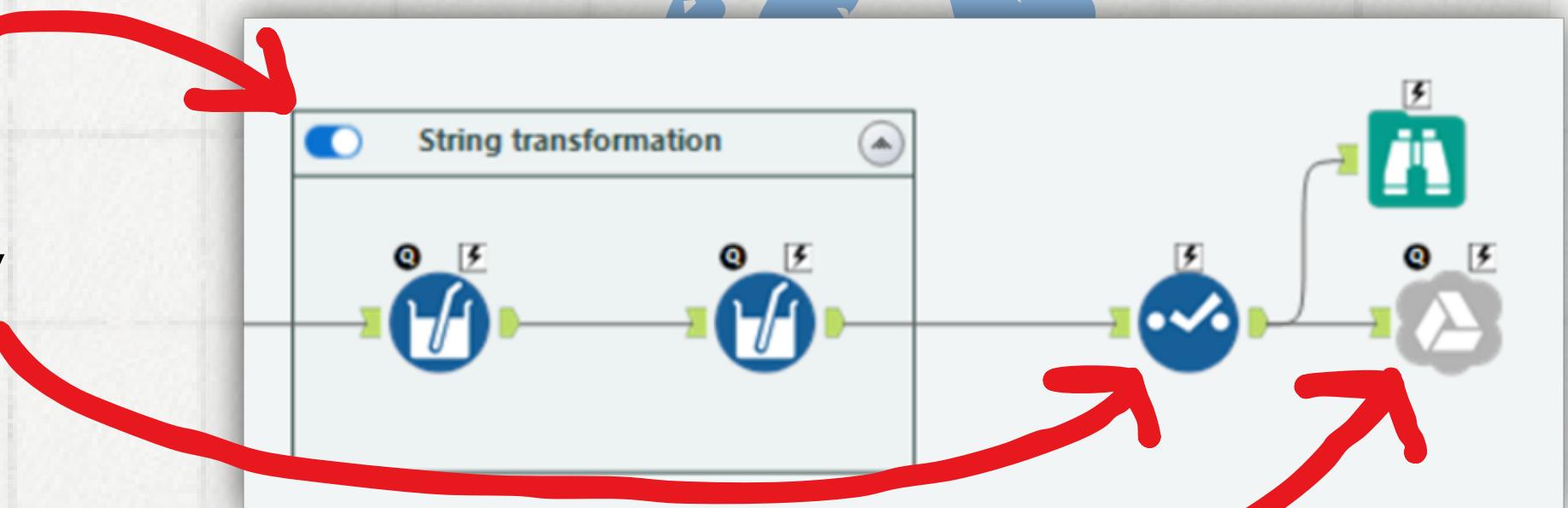
# Data display

- Replacing abbreviations
  - E.g. SD > Single (Single deck)
- Filter and cleanse data to display relevant details



# Data display

- Replacing abbreviations
  - E.g. SD > Single (Single deck)
- Filter and cleanse data to display relevant details
- Output to Google Sheets



# Data display

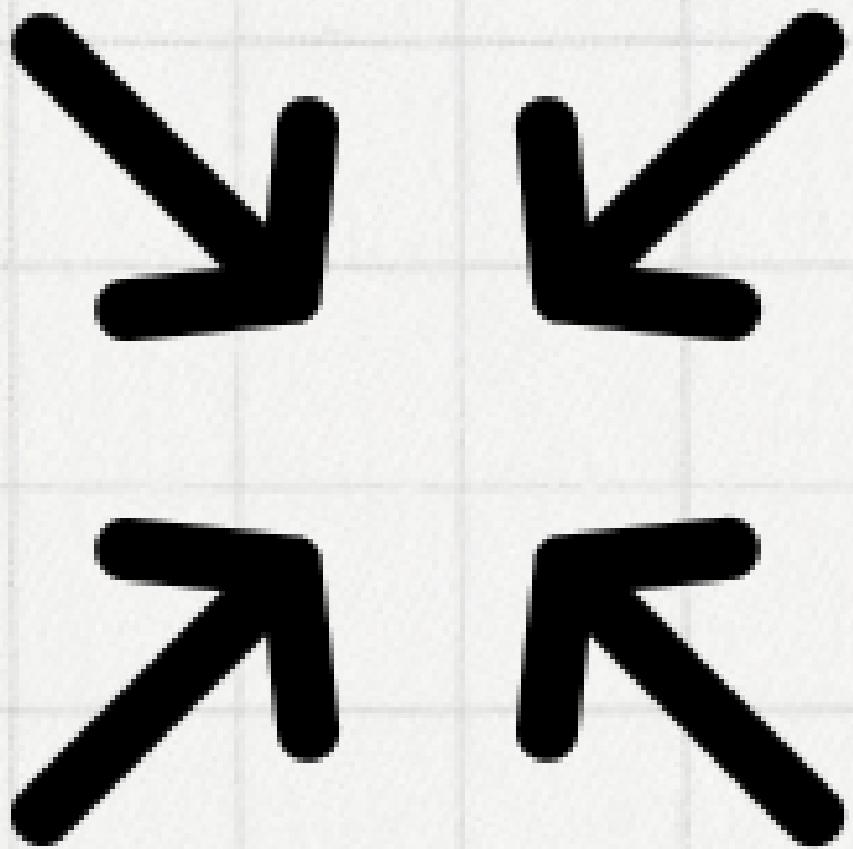
- Retrieves data from Google Sheets
- Keypad-like selection for bus services
- Important information for next 3 buses



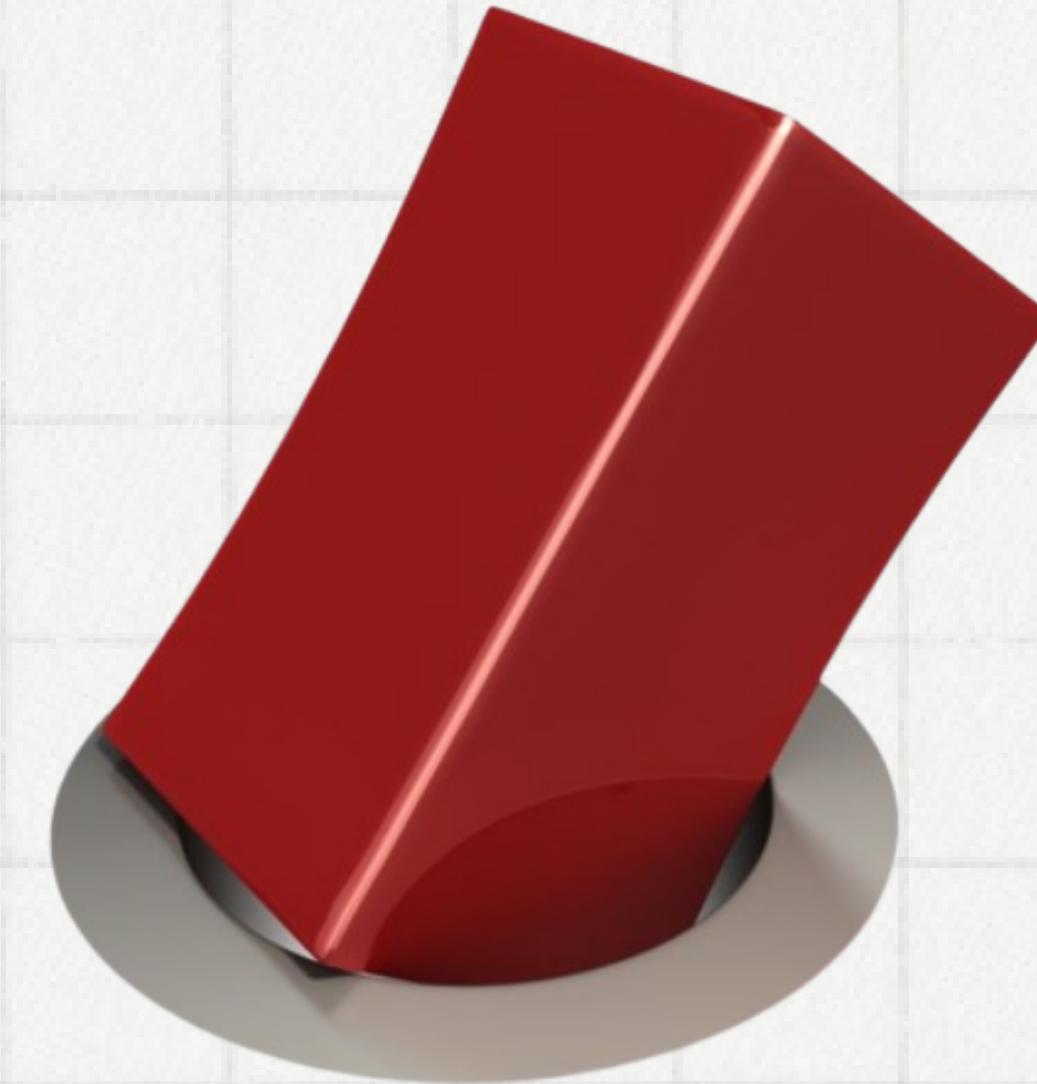
# **Findings**

## **Data & Alteryx**

# Data

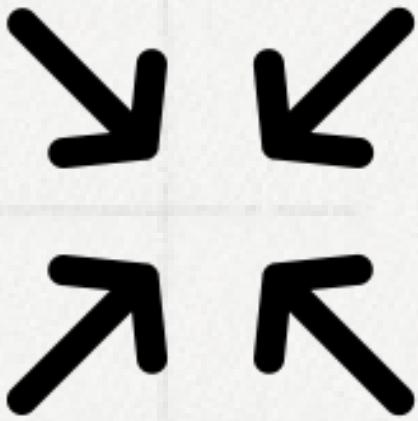


API Response shrinks



Improper data types

# Data

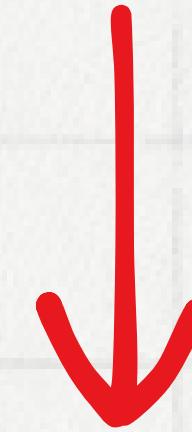


API Response shrinks

- As the last buses finish their shift, there is less data for the API to return.

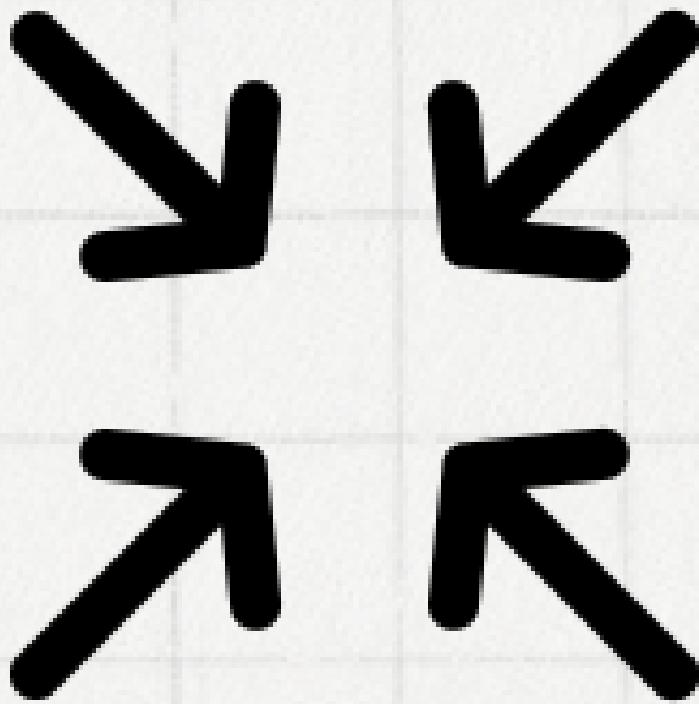
- Leads to null values
- Eventually, chunks of data disappear

A	B	C	D	E	F	G	H	I	J	K	L	M
Bus Number	1st Bus ETA	1st Bus Minutes	1st Bus Capacity	1st Bus Type	2nd Bus ETA	2nd Bus Minutes	2nd Bus Capacit	2nd Bus Type	3rd Bus ETA	3rd Bus Minutes	3rd Bus Capacity	3rd Bus Type
154	2024-02-02 17:1	13	High	Single	2024-02-02 17:1	13	High	Single	2024-02-02 17:1	15	High	Double
184	2024-02-02 17:0	6	Medium	Bendy	2024-02-02 17:1	11	High	Double	2024-02-02 17:2	18	High	Single
105	2024-02-02 17:0	3	High	Single	2024-02-02 17:1	9	High	Single	2024-02-02 17:2	21	High	Single
189	2024-02-02 17:0	0	High	Double	2024-02-02 17:1	8	High	Single	2024-02-02 17:1	15	High	Double
106	2024-02-02 17:0	0	Medium	Single	2024-02-02 17:1	14	High	Single	2024-02-02 17:2	16	High	Double
201	2024-02-02 17:0	5	High	Double	2024-02-02 17:1	16	High	Double	2024-02-02 17:3	27	Medium	Single
52	2024-02-02 17:0	3	High	Double	2024-02-02 17:0	6	High	Double	2024-02-02 17:2	18	High	Double
185	2024-02-02 17:0	4	High	Double	2024-02-02 17:1	15	High	Double	2024-02-02 17:2	23	High	Double
183	2024-02-02 17:0	0	High	Double	2024-02-02 17:1	8	High	Double	2024-02-02 17:2	17	High	Double



Bus Number	1st Bus ETA	1st Bus Minutes	1st Bus Capacity	1st Bus Type	2nd Bus ETA	2nd Bus Minutes	2nd Bus Capacit	2nd Bus Type	3rd Bus ETA	3rd Bus Minutes
201	2024-02-04 0:15	2	High	Double						
105	2024-02-04 0:18	5	High	Single						
154	2024-02-04 0:18	5	High	Double	2024-02-04 0:30	17	High	Double	2024-02-04 0:46	34

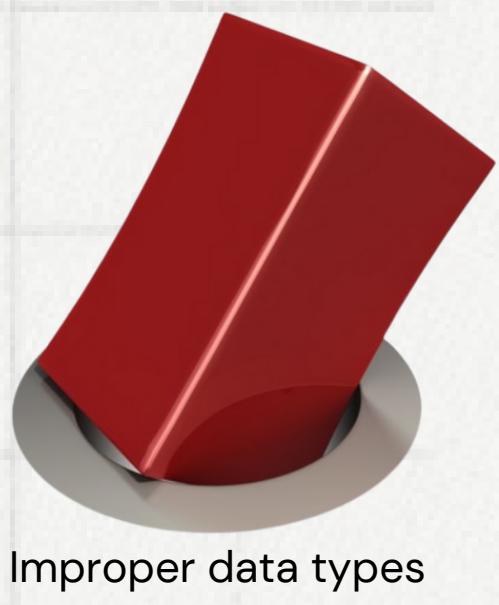
# Data



API Response shrinks

Workflow cannot be properly worked on or reviewed when it is approaching midnight and wee hours, unless data input is hardcoded into the workflow.

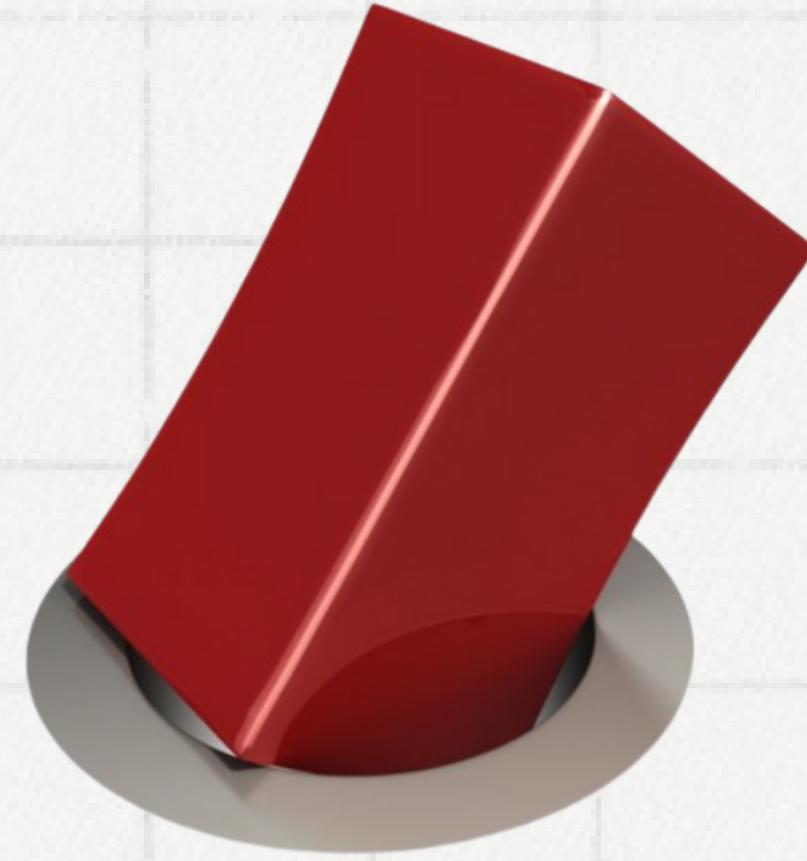
# Data



- Columns containing DateTime information not in recognizable DateTime format
- Manual conversion to DateTime, with specified custom format
- Separate tool processing for each of column
- Extract information like minutes to arrival



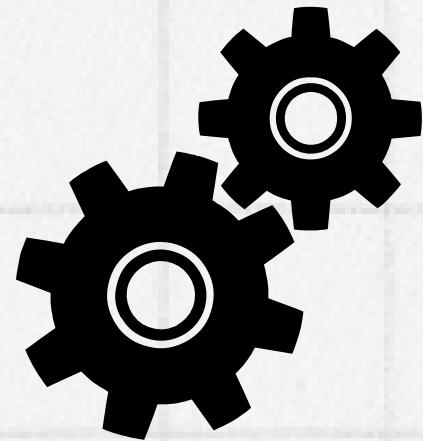
# Data



Improper data types

If API is enhanced in the future to provide details for more incoming buses, more tools will have to be included in the workflow, leading to increased computing loads and runtime. A workaround would be to sample data to only the next 3 buses.

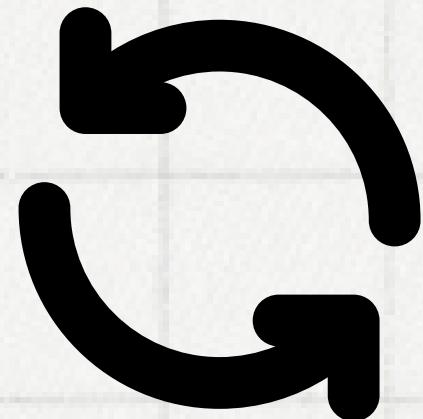
# Alteryx



Alteryx auto-initializes  
data types



Browse Tool slows down  
workflow

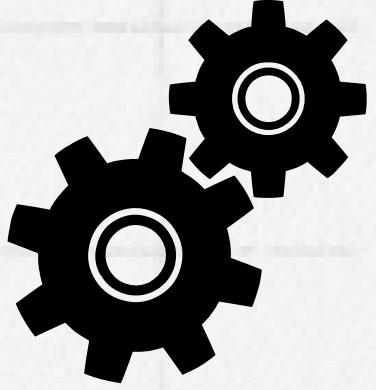


Tool configurations  
may reset



Alteryx supports  
additional packages

# Alteryx

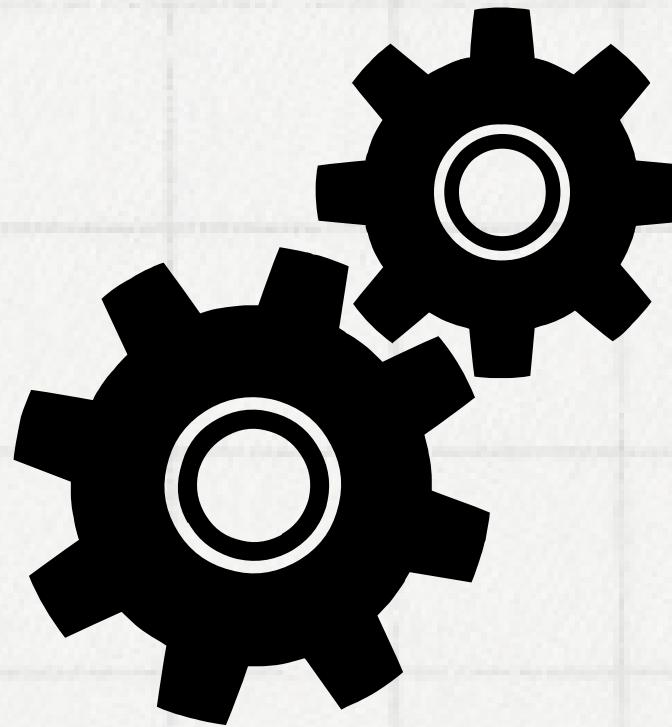


Alteryx auto-initializes  
data types

- When appropriate, Alteryx automatically converts data types.
  - E.g. a numerical column may have its data type changed to Int32 or Float
- Does not auto convert again when dealing with expressions

Parse Error at char(22): Type mismatch in operator +. (Expression #1)

# Alteryx



Alteryx auto-initializes  
data types

Before performing string expressions that involve columns with numbers, data types need to be checked beforehand to ensure no errors arise.

# Alteryx



Browse Tool slows down workflow

- Browse tool creates a .yxdb file
  - holds data fields and values of the data stream
- Creation of data takes time
- Allows for reviewing the effect of tools on data stream

JSON_Name	JSON_ValueString	1	2	3
Services.0.ServiceNo	105	Services	0	ServiceNo
Services.0.Operator	SBST	Services	0	Operator
Services.0.NextBus-OriginCode	28009	Services	0	NextBus-OriginCode
Services.0.NextBus-DestinationCode	66009	Services	0	NextBus-DestinationCode
Services.0.NextBus-EstimatedArrival	2024-02-01T17:48:30+08:00	Services	0	NextBus-EstimatedArrival
Services.0.NextBus-Latitude	1.3268023333333332	Services	0	NextBus-Latitude

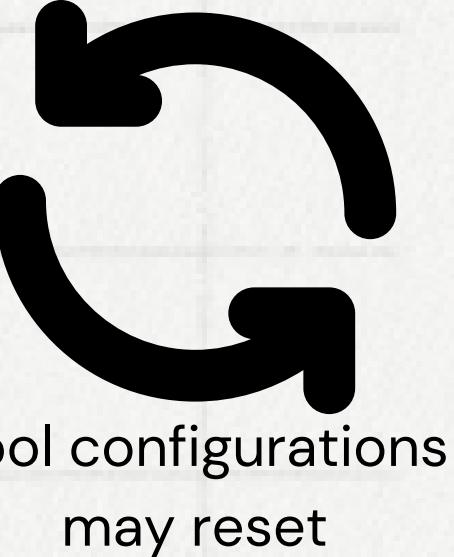
# Alteryx



Browse Tool slows down  
workflow

Since the tool has good benefits but increases computation time, it should be used in moderation.

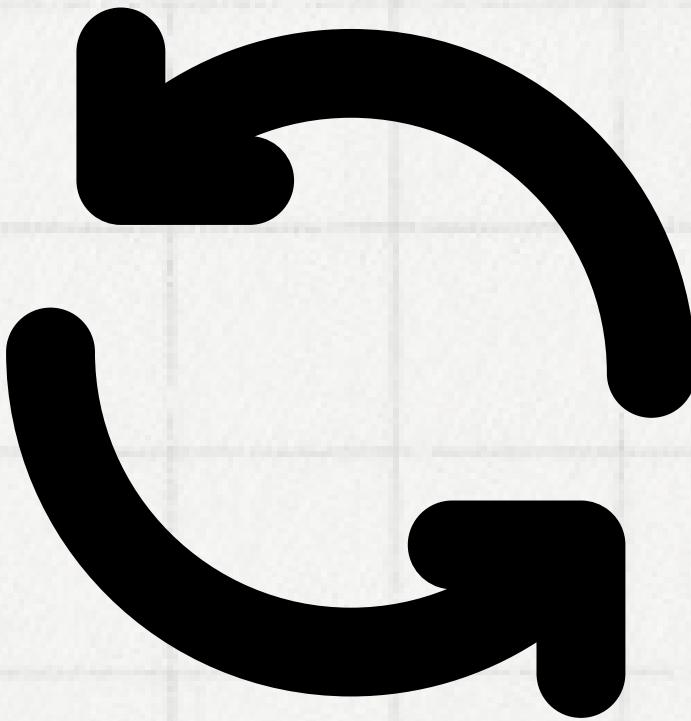
# Alteryx



Tool configurations  
may reset

- Most tools do not remember missing data
- When receiving blank data flow, raises an error and resets configuration
- Default configurations may lead to more errors, and more blank data flow downstream

# Alteryx



Tool configurations  
may reset

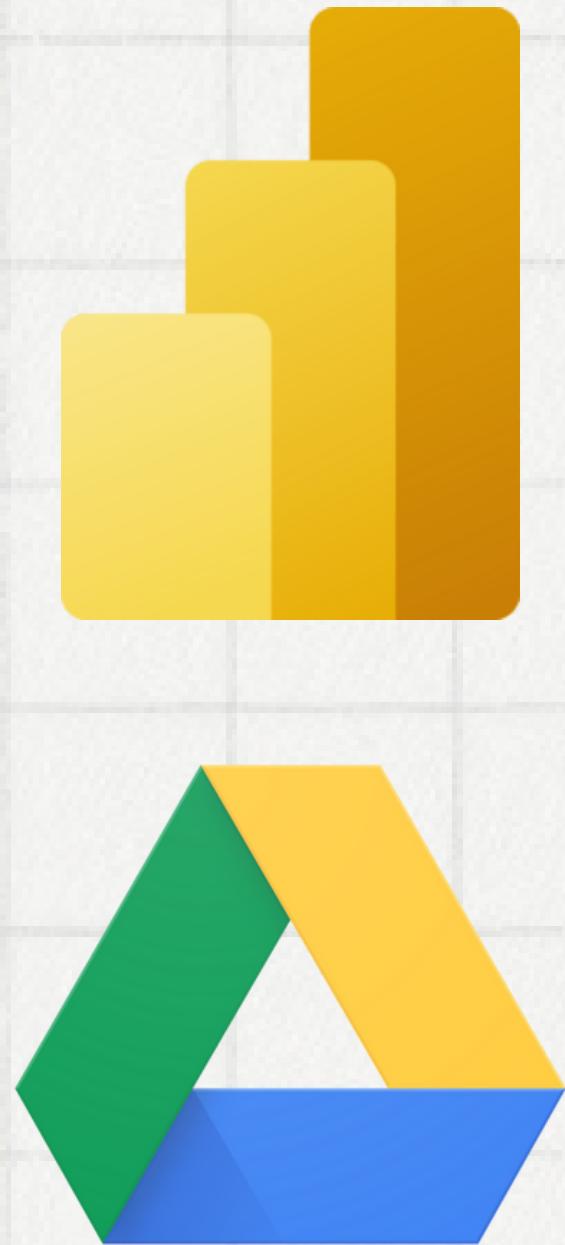
Extremely annoying and detrimental phenomenon that can be prevented by careful double-checking of the relationships between the data, workflow tools, and expressions.

# Alteryx



Alteryx supports  
additional packages

- Downloadable packages that support integration with other platforms, like Power BI & Google Drive

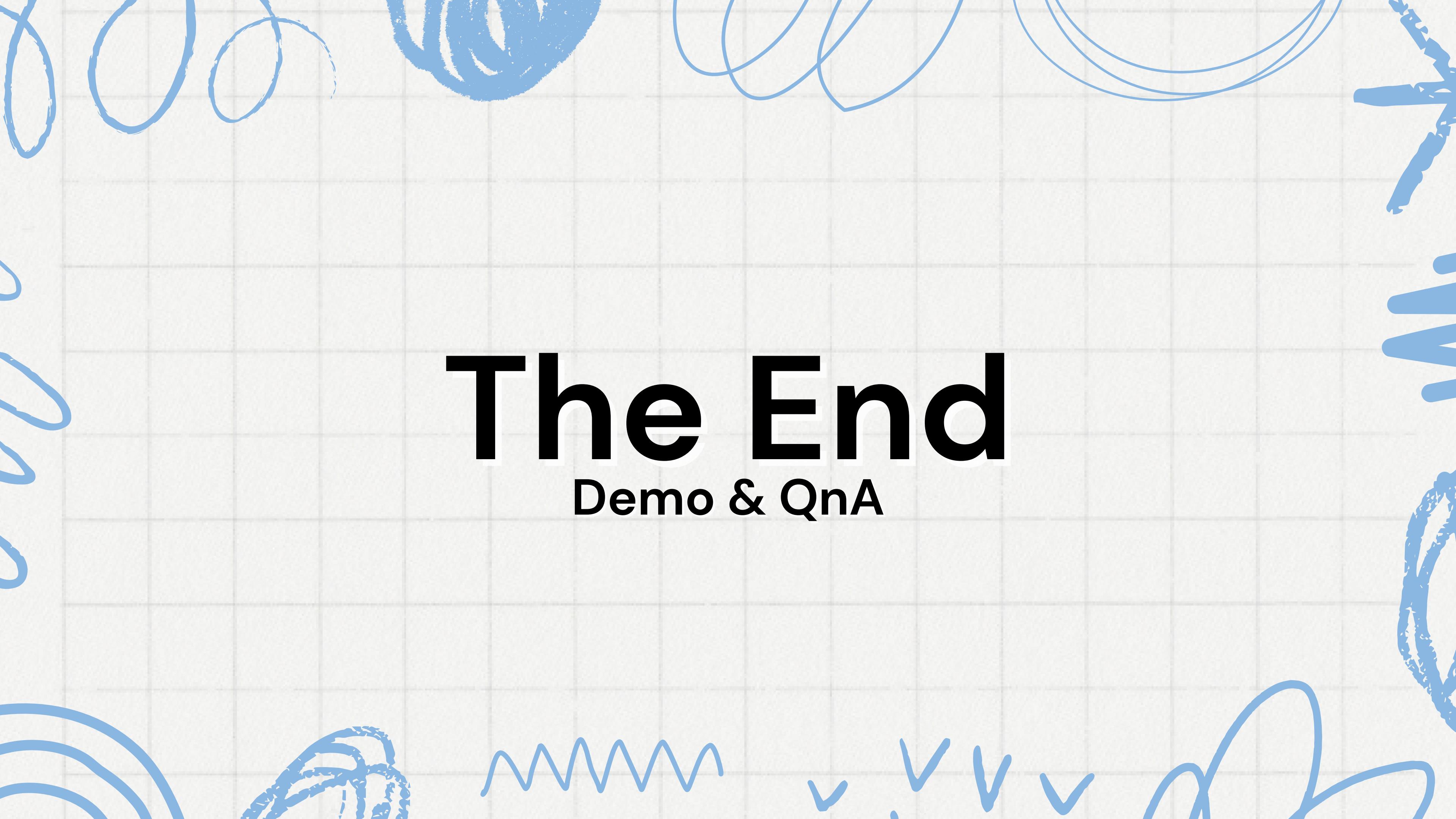


# Alteryx



Alteryx supports  
additional packages

Integration with other platforms can bring about benefits, e.g. connecting to Power BI can lead to enhanced reporting and visualization.



# **The End**

## **Demo & QnA**