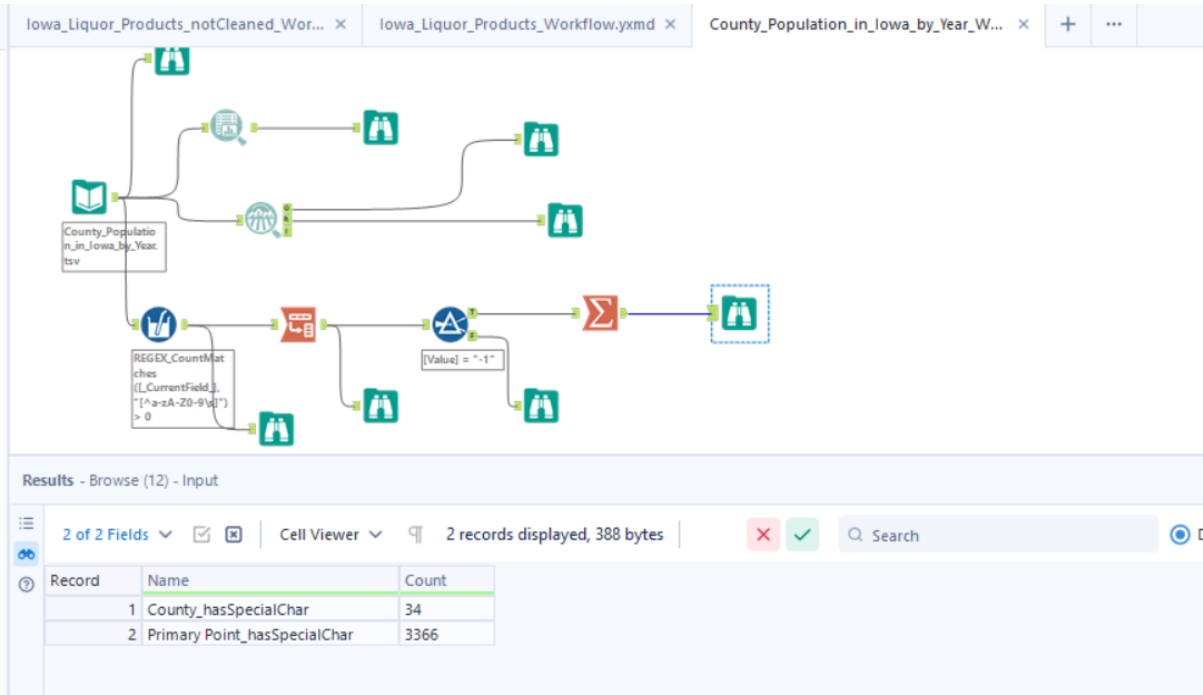
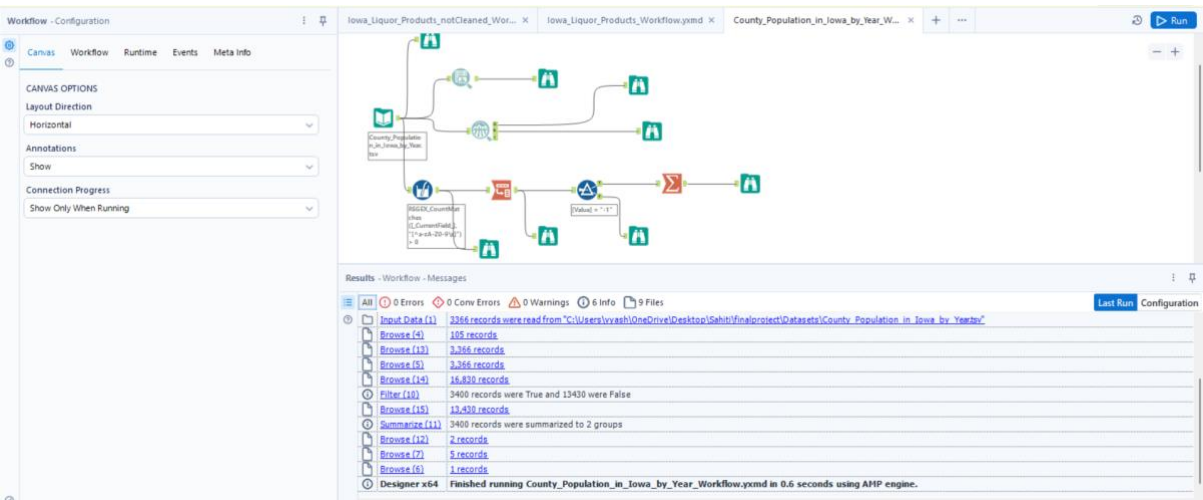


# COUNTY POPULATION BY YEAR



Workflow: Iowa\_Liquor\_Products\_Workflow.yxmd

Results - Browse (4) - Input

Record	FieldName	Name	Value
1	FIPS	Name	FIPS
2	FIPS	Data Type	V_WString
3	FIPS	Size	254
4	FIPS	Source	File: C:\Users\vyash\OneDrive\Desktop\Sahiti\fin...
5	FIPS	Description	[Null]
6	FIPS	OKs	3366
7	FIPS	Nulls	0
8	FIPS	Non-Nulls	3366
9	FIPS	Blanks	0
10	FIPS	Values with Leading Whitespace	0
11	FIPS	Values with Trailing Whitespace	0

Inference – There are 3366 records, with PRIMARY POINT that needs to be split into LATITUTDE AND LONGITUDE

## IOWA\_LIQUOR\_PRODUCTS

Workflow: Iowa\_Liquor\_Products\_Workflow.yxmd

Results - Workflow - Messages

- Designer x64** The Designer x64 reported: Allocating requested memory would be more than available physical memory. Reverting to 5675.9 MB of memory.
- Designer x64** The Designer x64 reported: This is AMP Engine; running 16 worker threads; memory limit 5675.9 MB.
- Input Data (1)** Autodetected Code Page: UTF-8
- Input Data (1)** 8,776 records were read from "C:\Users\vyash\OneDrive\Desktop\Sahiti\Finalproject\Datasets\Iowa\_Liquor\_Products.tsv"
- Browse (5)** 8,776 records
- Browse (4)** 357 records
- Browse (12)** 8,776 records
- Browse (6)** 149,192 records
- Filter (7)** 49622 records were True and 99570 were False
- Browse (8)** 49,622 records
- Summarize (9)** 49622 records were summarized to 8 groups
- Browse (10)** 8 records
- Designer x64** Finished running Iowa\_Liquor\_Products\_Workflow.yxmd in 0.6 seconds using AMP engine.

Workflow: iowa\_Liquor\_Products\_Workflow.yxmd

REGEX\_CountMatches configuration:  
 ([CurrentField\_1])  
 ["^a-zA-Z0-9a"]  
 > 0

Results - Browse (10) - Input

2 of 2 Fields | Cell Viewer | 8 records displayed, 509 bytes | Search | Data | Metadata

Record	Name	Count
1	Category Name_hasSpecialChar	2657
2	State Case Cost_hasSpecialChar	8776
3	Item Description_hasSpecialChar	948
4	List Date_hasSpecialChar	8776
5	Vendor Name_hasSpecialChar	2137
6	State Bottle Cost_hasSpecialChar	8776
7	State Bottle Retail_hasSpecialChar	8776
8	Report Date_hasSpecialChar	8776

Workflow: iowa\_Liquor\_Products\_Workflow.yxmd

REGEX\_CountMatches configuration:  
 ([CurrentField\_1])  
 ["^a-zA-Z0-9a"]  
 > 0

Results - Browse (4) - Input

3 of 3 Fields | Cell Viewer | 357 records displayed, 5,309 bytes | Search | Data | Metadata | Actions

Record	FieldName	Name	Value
1	Item Number	Name	Item Number
2	Item Number	Data Type	V_WString
3	Item Number	Size	254
4	Item Number	Source	File: C:\Users\vyash\OneDrive\Desktop\Sahiti\fin...
5	Item Number	Description	[Null]
6	Item Number	OKs	8776
7	Item Number	Nulls	0
8	Item Number	Non-Nulls	8776
9	Item Number	Blanks	0
10	Item Number	Values with Leading Whitespace	0
11	Item Number	Values with Trailing Whitespace	0

Inference – This has 8776 records with all camel case that have been converted to upper case while cleaning. There are records with UPC corresponding to NULL values in SCC.

## IOWA\_LIQUOR\_PRODUCTS\_NOTCLEANED

Iowa\_Liquor\_Products\_notCleaned\_Wor... x +

The workflow diagram shows a sequence of steps: 1. Input Data (1) - Autodetected Code Page: UTF-8. 2. Browse (4) - 6,350 records. 3. Browse (3) - 357 records. 4. Browse (11) - 6,350 records. 5. Browse (5) - 107,950 records. 6. Filter (6) - 35533 records were True and 72417 were False. 7. Browse (7) - 35,533 records. 8. Summarize (8) - 35533 records were summarized to 8 groups. 9. Browse (9) - 8 records. The workflow is executed using the AMP engine.

Results - Workflow - Messages

All 0 Errors 0 Conv Errors 0 Warnings 6 Info 7 Files Last Run C

Designer x64 The Designer x64 reported: This is AMP Engine; running 16 worker threads; memory limit 5572.1 MB.

Input Data (1) Autodetected Code Page: UTF-8

Input Data (1) 6350 records were read from "C:\Users\vvash\OneDrive\Desktop\Sahiti\Finalproject\Datasets\Iowa\_Liquor\_Products\_notCleaned.csv"

Browse (4) 6,350 records

Browse (3) 357 records

Browse (11) 6,350 records

Browse (5) 107,950 records

Filter (6) 35533 records were True and 72417 were False

Browse (7) 35,533 records

Summarize (8) 35533 records were summarized to 8 groups

Browse (9) 8 records

Designer x64 Finished running Iowa\_Liquor\_Products\_notCleaned\_Workflow.yxmd in 0.6 seconds using AMP engine.

Iowa\_Liquor\_Products\_notCleaned\_Wor... x +

The workflow diagram shows a sequence of steps: 1. Input Data (1) - Autodetected Code Page: UTF-8. 2. Browse (4) - 6,350 records. 3. Browse (3) - 357 records. 4. Browse (11) - 6,350 records. 5. Browse (5) - 107,950 records. 6. Filter (6) - 35533 records were True and 72417 were False. 7. Browse (7) - 35,533 records. 8. Summarize (8) - 35533 records were summarized to 8 groups. 9. Browse (9) - 8 records. The workflow is executed using the AMP engine.

Results - Browse (9) - Input

2 of 2 Fields Cell Viewer 8 records displayed, 509 bytes Search

Record	Name	Count
1	Category Name_hasSpecialChar	1733
2	State Case Cost_hasSpecialChar	6350
3	Item Description_hasSpecialChar	633
4	List Date_hasSpecialChar	6350
5	Vendor Name_hasSpecialChar	1417
6	State Bottle Cost_hasSpecialChar	6350
7	State Bottle Retail_hasSpecialChar	6350
8	Report Date_hasSpecialChar	6350

lowa\_Liquor\_Products\_notCleaned\_Wor... x +

```

Source
    > Filter
    > Group By
    > Pivot
    > Unpivot
    > Append
  
```

REGEX\_CountMatches([CurrentField], "[^a-zA-Z0-9/s]") > 0

[Value] = "-1"

Results - Browse (3) - Input

3 of 3 Fields | Cell Viewer | 357 records displayed, 6,378 bytes | Search

Record	FieldName	Name	Value
1	Item Number	Name	Item Number
2	Item Number	Data Type	V_WString
3	Item Number	Size	254
4	Item Number	Source	File: C:\Users\vyash\OneDrive\Desktop\Sahiti\fin...
5	Item Number	Description	[Null]
6	Item Number	OKs	6350
7	Item Number	Nulls	0
8	Item Number	Non-Nulls	6350
9	Item Number	Blanks	0
10	Item Number	Values with Leading Whitespace	0
11	Item Number	Values with Trailing Whitespace	0

Inference – There are 6350 records with camel cases. There are fields with SCC values corresponding to NULL values in UPC field.

Difference between CLEANED AND NOTCLEANED files – data value differences between UPC and SCC as mentioned above and there are 6344 records common between both the files i.e. there are 6 records present in NOTCLEANED file which are missing from CLEANED, we plan to deal with this by merging both the files hence getting all records.

# IOWA\_LIQUOR\_STORES

Workflow visualization for Iowa\_Liquor\_Sales\_Workflow.yxmd and Iowa\_Liquor\_Stores\_Workflow.yxmd. The workflow starts with an input data source (Iowa\_Liquor\_Stores.tsv) and proceeds through several steps: a filter, a join, a filter, a join, a filter, a join, and a final output. The workflow is executed using the AMP engine.

Results - Workflow - Messages

All 0 Errors 0 Conv Errors 0 Warnings 6 Info 7 Files

Designer x64 The Designer x64 reported: This is AMP Engine; running 16 worker threads; memory limit 5639.4 MB.

Input Data (1) Autodetected CodePage: UTF-8

Input Data (1) 2626 records were read from "C:\Users\vyvash\OneDrive\Desktop\Sahiti\finalproject\Datasets\Iowa\_Liquor\_Stores.tsv"

Browse (4) 2,626 records

Browse (3) 189 records

Browse (12) 2,626 records

Browse (17) 23,634 records

Filter (13) 7893 records were True and 15741 were False

Browse (18) 7,893 records

Summarize (15) 7893 records were summarized to 5 groups

Browse (16) 5 records

Designer x64 Finished running Iowa\_Liquor\_Stores\_Workflow.yxmd in 0.5 seconds using AMP engine.

Workflow visualization for Iowa\_Liquor\_Stores\_Workflow.yxmd. The workflow starts with an input data source (Iowa\_Liquor\_Stores.tsv) and proceeds through several steps: a filter, a join, a filter, a join, a filter, a join, and a final output. The workflow is executed using the AMP engine.

Results - Browse (16) - Input

2 of 2 Fields Cell Viewer 5 records displayed, 433 bytes

Record	Name	Count
1	Address_hasSpecialChar	341
2	City_hasSpecialChar	3
3	Store Address_hasSpecialChar	2626
4	Report Date_hasSpecialChar	2626
5	Name_hasSpecialChar	2297

Workflow: Iowa\_Liquor\_Sales\_Workflow.yxmd x Iowa\_Liquor\_Stores\_Workflow.yxmd\*

Results - Browse (3) - Input

3 of 3 Fields | Cell Viewer | 189 records displayed, 2,625 bytes | Search | Data | Metadata | Actions | 00

Record	FieldName	Name	Value
1	Store	Name	Store
2	Store	Data Type	V_WString
3	Store	Size	254
4	Store	Source	File: C:\Users\vyash\OneDrive\Desktop\Sahiti\fin...
5	Store	Description	[Null]
6	Store	OKs	2626
7	Store	Nulls	0
8	Store	Non-Nulls	2626
9	Store	Blanks	0
10	Store	Values with Leading Whitespace	0
11	Store	Values with Trailing Whitespace	0

Inference – There are 2626 records with no NULLS.

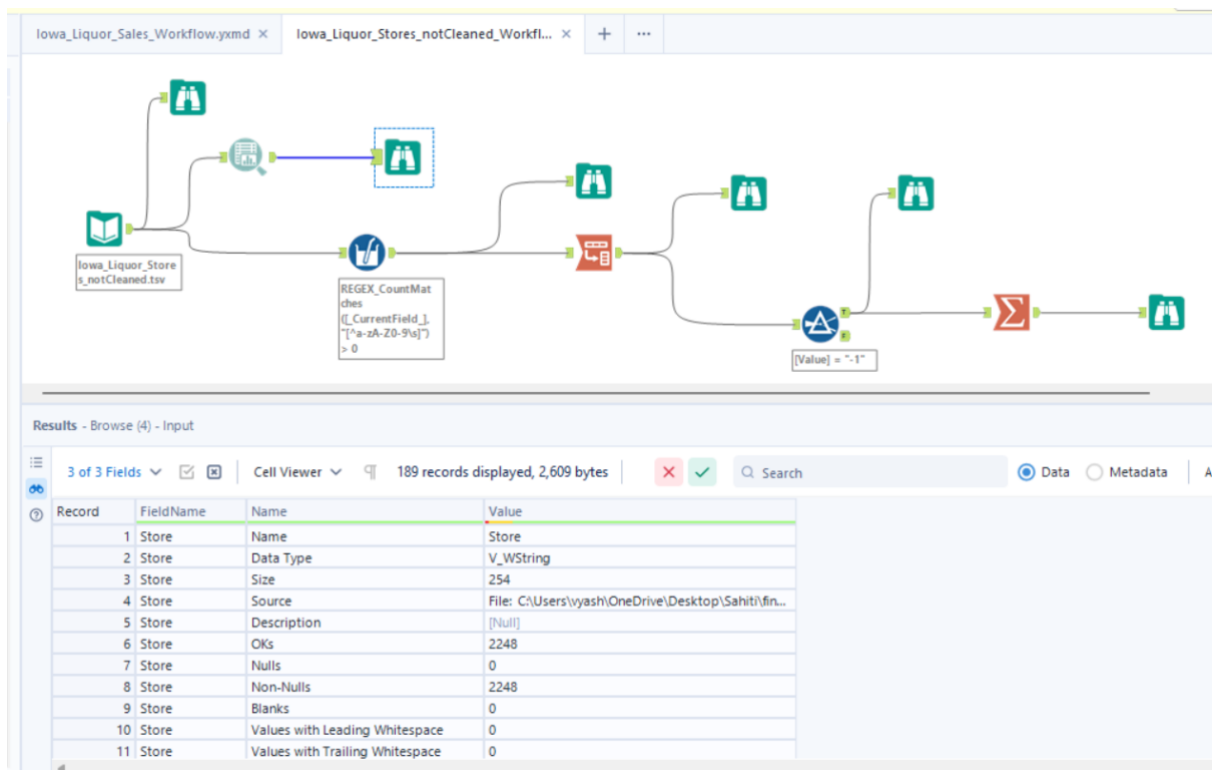
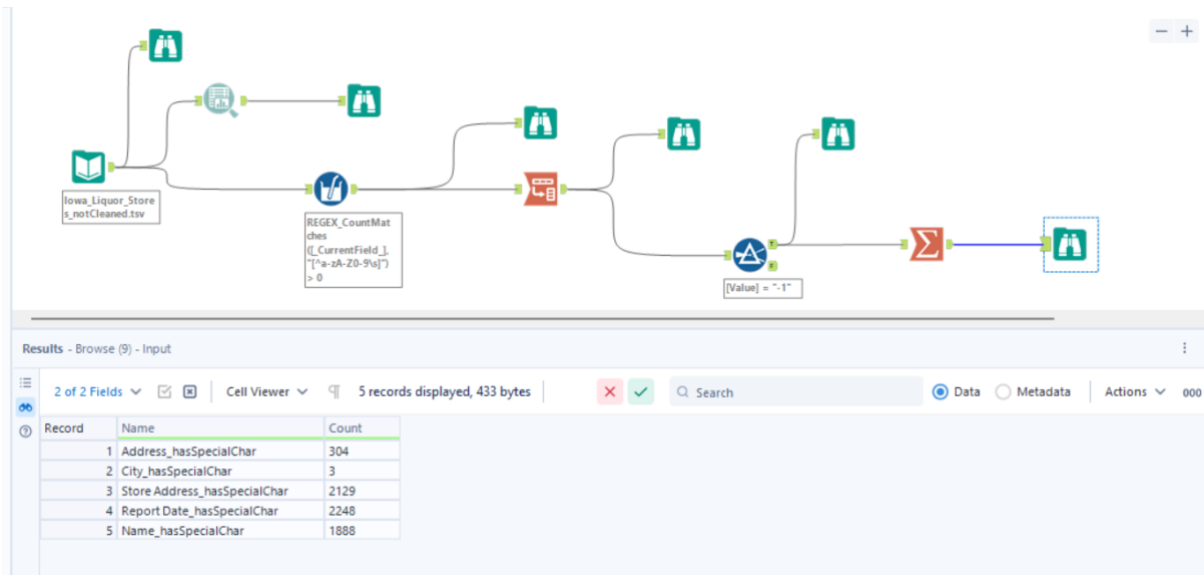
## IOWA\_LIQUOR\_STORES\_NOTCLEANED

Workflow: Iowa\_Liquor\_Sales\_Workflow.yxmd x Iowa\_Liquor\_Stores\_notCleaned\_Workfl...

Results - Workflow - Messages

All | 0 Errors | 0 Conv Errors | 0 Warnings | 6 Info | 7 Files | Last Run | Configuration

- Designer x64 The Designer x64 reported: Allocating requested memory would be more than available physical memory. Reverting to 5435.0 MB of memory.
- Designer x64 The Designer x64 reported: This is AMP Engine; running 16 worker threads; memory limit 5435.0 MB.
- Input Data (2) Autodetected Code Page: UTF-8
- Input Data (2) 2248 records were read from "C:\Users\vyash\OneDrive\Desktop\Sahiti\finalproject\Datasets\Iowa\_Liquor\_Stores\_notCleaned.csv"
- Browse (1) 2,248 records
- Browse (4) 189 records
- Browse (11) 2,248 records
- Browse (5) 20,232 records
- Filter (6) 6572 records were True and 13660 were False
- Browse (7) 6,572 records
- Summarize (8) 6572 records were summarized to 5 groups
- Browse (9) 5 records
- Designer x64 Finished running Iowa\_Liquor\_Stores\_notCleaned\_Workflow.yxmd in 0.4 seconds using AMP engine.




Inference – There are 2248 records with NULLs in Address field.

Differences – Except for the NULLs there are not many differences. NOTCLEANED and CLEANED files have 2248 common records i.e. all records from NOTCLEANED file are present in CLEANED file, hence no missing records.



IOWA\_LIQUOR\_SALES\_NOTCLEANED

Iowa\_Liquor\_Sales\_Workflow.yxmd xIowa\_Liquor\_Sales\_notCleaned\_Workflo... x+...



Results - Workflow - Messages

All0 Errors0 Conv Errors0 Warnings6 Info7 Files

Designer x64

The Designer x64 reported: This is AMP Engine; running 16 worker threads; memory limit 6400.0 MB.

Input Data (1)

Autodetected Code Page: UTF-8

Designer x64

The Designer x64 reported: Beginning to compact waiting packets to reduce memory usage

Input Data (1)

24847481 records were read from "C:\Users\vyash\OneDrive\Desktop\Sahiti\Finalproject\Datasets\Iowa\_Liquor\_Sales\_notCleaned.csv"

Browse (4)

24,847,481 records

Browse (11)

24,847,481 records

Browse (5)

99,389,924 records

Filter (6)

74605657 records were True and 24784267 were False

Browse (7)

74,605,657 records

Summarize (8)

74605657 records were summarized to 4 groups

Browse (9)

4 records


Browse (3)

504 records

Designer x64

Finished running Iowa\_Liquor\_Sales\_notCleaned\_Workflow.yxmd in 4:57 minutes using AMP engine.

Iowa\_Liquor\_Sales\_Workflow.yxmd xIowa\_Liquor\_Sales\_notCleaned\_Workflo... x+...



Results - Browse (9) - Input

2 of 2 FieldsCell Viewer4 records displayed, 444 bytes

Record

Name

Count

1

County\_hasSpecialChar

63224

2

Sale (Dollars)\_hasSpecialChar

24847471

3

Date\_hasSpecialChar

24847481

4

Volume Sold (Gallons)\_hasSpecialChar

24847481

## IOWA\_LIQUOR\_SALES\_BY\_YEAR

Iowa\_Liquor\_Sales\_Workflow.yxmd x Iowa\_Liquor\_Sales\_by\_Year\_Workflow.yx... x + ...

The workflow diagram shows a sequence of steps: 1. Input Data (1) - Autodetected Code Page: UTF-8. 2. Filter (6) - 26 records were True and 13 were False. 3. Summarize (8) - 26 records were summarized to 2 groups. 4. Browse (3) - 63 records. 5. Browse (4) - 13 records. 6. Browse (7) - 26 records. 7. Browse (11) - 13 records. 8. Browse (9) - 2 records. 9. Browse (5) - 39 records. 10. Designer x64 - Finished running Iowa\_Liquor\_Sales\_by\_Year\_Workflow.yxmd in 0.4 seconds using AMP engine.

Results - Workflow - Messages

All 0 Errors 0 Conv Errors 0 Warnings 6 Info 7 Files

Designer x64 The Designer x64 reported: This is AMP Engine; running 16 worker threads; memory limit 5518.6 MB.

Input Data (1) Autodetected Code Page: UTF-8

Input Data (1) 13 records were read from "C:\Users\vyash\OneDrive\Desktop\Sahiti\finalproject\Datasets\Iowa\_Liquor\_Sales\_by\_Year.csv"

Filter (6) 26 records were True and 13 were False

Summarize (8) 26 records were summarized to 2 groups

Browse (3) 63 records

Browse (4) 13 records

Browse (7) 26 records

Browse (11) 13 records

Browse (9) 2 records

Browse (5) 39 records

Designer x64 Finished running Iowa\_Liquor\_Sales\_by\_Year\_Workflow.yxmd in 0.4 seconds using AMP engine.

Iowa\_Liquor\_Sales\_Workflow.yxmd x Iowa\_Liquor\_Sales\_by\_Year\_Workflow.yx... x + ...

The workflow diagram shows a sequence of steps: 1. Input Data (1) - Autodetected Code Page: UTF-8. 2. Filter (6) - 26 records were True and 13 were False. 3. Summarize (8) - 26 records were summarized to 2 groups. 4. Browse (3) - 63 records. 5. Browse (4) - 13 records. 6. Browse (7) - 26 records. 7. Browse (11) - 13 records. 8. Browse (9) - 2 records. 9. Browse (5) - 39 records. 10. Designer x64 - Finished running Iowa\_Liquor\_Sales\_by\_Year\_Workflow.yxmd in 0.4 seconds using AMP engine.

Results - Browse (9) - Input

2 of 2 Fields 2 records displayed, 398 bytes

Record	Name	Count
1	Sale (Dollars)_hasSpecialChar	13
2	Volume Sold (Gallons)_hasSpecialChar	13

Iowa\_Liquor\_Sales\_Workflow.yxmd x Iowa\_Liquor\_Sales\_by\_Year\_Workflow.yx...

Results - Browse (3) - Input

3 of 3 Fields | Cell Viewer | 63 records displayed, 1,580 bytes | Search

Record	FieldName	Name	Value
1	Date	Name	Date
2	Date	Data Type	V_WString
3	Date	Size	254
4	Date	Source	File: C:\Users\vyash\OneDrive\Desktop\Sahiti\fin...
5	Date	Description	[Null]
6	Date	OKs	13
7	Date	Nulls	0
8	Date	Non-Nulls	13
9	Date	Blanks	0
10	Date	Values with Leading Whitespace	0
11	Date	Values with Trailing Whitespace	0

Inference – There are 13 records. It gives the sales for year wise.

## IOWA LIQUOR SALES BY YEAR COUNTY

Iowa\_Liquor\_Sales\_Workflow.yxmd x Iowa\_Liquor\_Sales\_by\_Year\_and\_County...

Results - Workflow - Messages

All | 0 Errors | 0 Conv Errors | 0 Warnings | 6 Info | 7 Files

- Designer x64** The Designer x64 reported: This is AMP Engine; running 16 worker threads; memory limit 5477.4 MB.
- Input Data (1)** Autodetected Code Page: UTF-8
- Input Data (1)** 1301 records were read from "C:\Users\vyash\OneDrive\Desktop\Sahiti\finalproject\Datasets\Iowa Liquor Sales by Year and County.csv"
- Browse (4)** 1,301 records
- Browse (3)** 84 records
- Browse (11)** 1,301 records
- Browse (5)** 5,204 records
- Filter (6)** 2615 records were True and 2589 were False
- Browse (7)** 2,615 records
- Summarize (8)** 2615 records were summarized to 3 groups
- Browse (9)** 3 records
- Designer x64** Finished running Iowa\_Liquor\_Sales\_by\_Year\_and\_County\_Workflow.yxmd in 0.5 seconds using AMP engine.

Iowa\_Liquor\_Sales\_Workflow.yxmd x Iowa\_Liquor\_Sales\_by\_Year\_and\_County... x + ...

The workflow starts with an input file 'Iowa\_Liquor\_Sales\_by\_Year\_and\_County.tsv'. It passes through a 'Filter' tool with the configuration: 'REGEX\_CountMatches ([CurrentField\_], "[^a-zA-Z0-9\s]\*") > 0'. The filtered data then goes through a 'Join' tool, followed by an 'Aggregate' tool with the configuration '[Value] = ~1\*'. The final output is a table with 3 records.

Results - Browse (9) - Input

2 of 2 Fields | Cell Viewer | 3 records displayed, 415 bytes | Search

Record	Name	Count
1	County_hasSpecialChar	13
2	Sale (Dollars)_hasSpecialChar	1301
3	Volume Sold (Gallons)_hasSpecialChar	1301

Iowa\_Liquor\_Sales\_Workflow.yxmd x Iowa\_Liquor\_Sales\_by\_Year\_and\_County... x + ...

The workflow starts with an input file 'Iowa\_Liquor\_Sales\_by\_Year\_and\_County.tsv'. It passes through a 'Filter' tool with the configuration: 'REGEX\_CountMatches ([CurrentField\_], "[^a-zA-Z0-9\s]\*") > 0'. The filtered data then goes through a 'Join' tool, followed by an 'Aggregate' tool with the configuration '[Value] = ~1\*'. The final output is a table with 84 records.

Results - Browse (3) - Input

3 of 3 Fields | Cell Viewer | 84 records displayed, 2,268 bytes | Search

Record	FieldName	Name	Value
1	Date	Name	Date
2	Date	Data Type	V_WString
3	Date	Size	254
4	Date	Source	File: C:\Users\vyash\OneDrive\Desktop\Sahiti\fin...
5	Date	Description	[Null]
6	Date	OKs	1301
7	Date	Nulls	0
8	Date	Non-Nulls	1301
9	Date	Blanks	0
10	Date	Values with Leading Whitespace	0

Inference – There are 1301 records where DATE and COUNTY fields act as a combined primary key. It gives the sales numbers by year and county wise.

## CITY\_POPULATION\_BY\_YEAR

Workflow diagram showing the process of calculating city population by year. The workflow starts with an input data source, followed by a filter step (REGEX\_CountMatches) and a join step. The final output is a summary table.

Results - Workflow - Messages

All 0 Errors 0 Conv Errors 0 Warnings 6 Info 7 Files Last Run

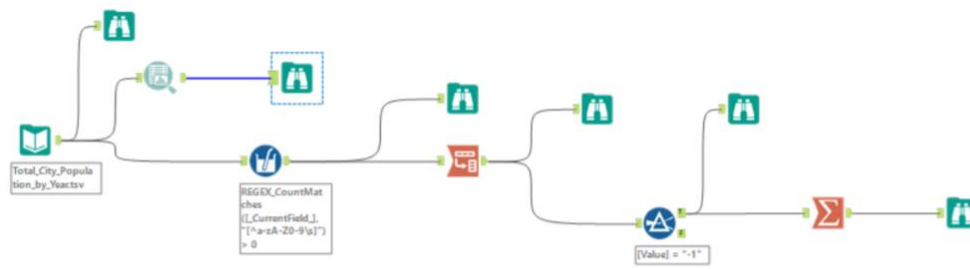
- Designer x64 The Designer x64 reported: Allocating requested memory would be more than available physical memory. Reverting to 5610.4 MB of memory.
- Designer x64 The Designer x64 reported: This is AMP Engine; running 16 worker threads; memory limit 5610.4 MB.
- Input Data (2) Autodetected Code Page: UTF-8
- Input Data (2) 32,152 records were read from "C:\Users\vyash\OneDrive\Desktop\Sahil\finalproject\Datasets\Total\_City\_Population\_by\_Year.csv"
- Browse (1) 32,152 records
- Browse (4) 105 records
- Browse (11) 32,152 records
- Browse (5) 160,760 records
- Filter (6) 32,282 records were True and 128,478 were False
- Browse (7) 32,282 records
- Summarize (8) 32,282 records were summarized to 2 groups
- Browse (9) 2 records
- Designer x64 Finished running Total\_City\_Population\_by\_Year\_Workflow.yxmd in 0.7 seconds using AMP engine.

Workflow diagram showing the process of calculating city population by year. The workflow starts with an input data source, followed by a filter step (REGEX\_CountMatches) and a join step. The final output is a summary table.

Results - Browse (9) - Input

2 of 2 Fields ☒ ☒ Cell Viewer ☒ 2 records displayed, 385 bytes ✖ ✓  🔍

Record	Name	Count
1	Primary Point_hasSpecialChar	31976
2	City_hasSpecialChar	306



Results - Browse (4) - Input

3 of 3 Fields | Cell Viewer | 105 records displayed, 1,905 bytes | Search | Data | Metadata

Record	FieldName	Name	Value
1	FIPS	Name	FIPS
2	FIPS	Data Type	V_WString
3	FIPS	Size	254
4	FIPS	Source	File: C:\Users\vyash\OneDrive\Desktop\Sahiti\fin...
5	FIPS	Description	[Null]
6	FIPS	OKs	32152
7	FIPS	Nulls	0
8	FIPS	Non-Nulls	32152
9	FIPS	Blanks	0
10	FIPS	Values with Leading Whitespace	0
11	FIPS	Values with Trailing Whitespace	0

Inference – There are 32152 records.