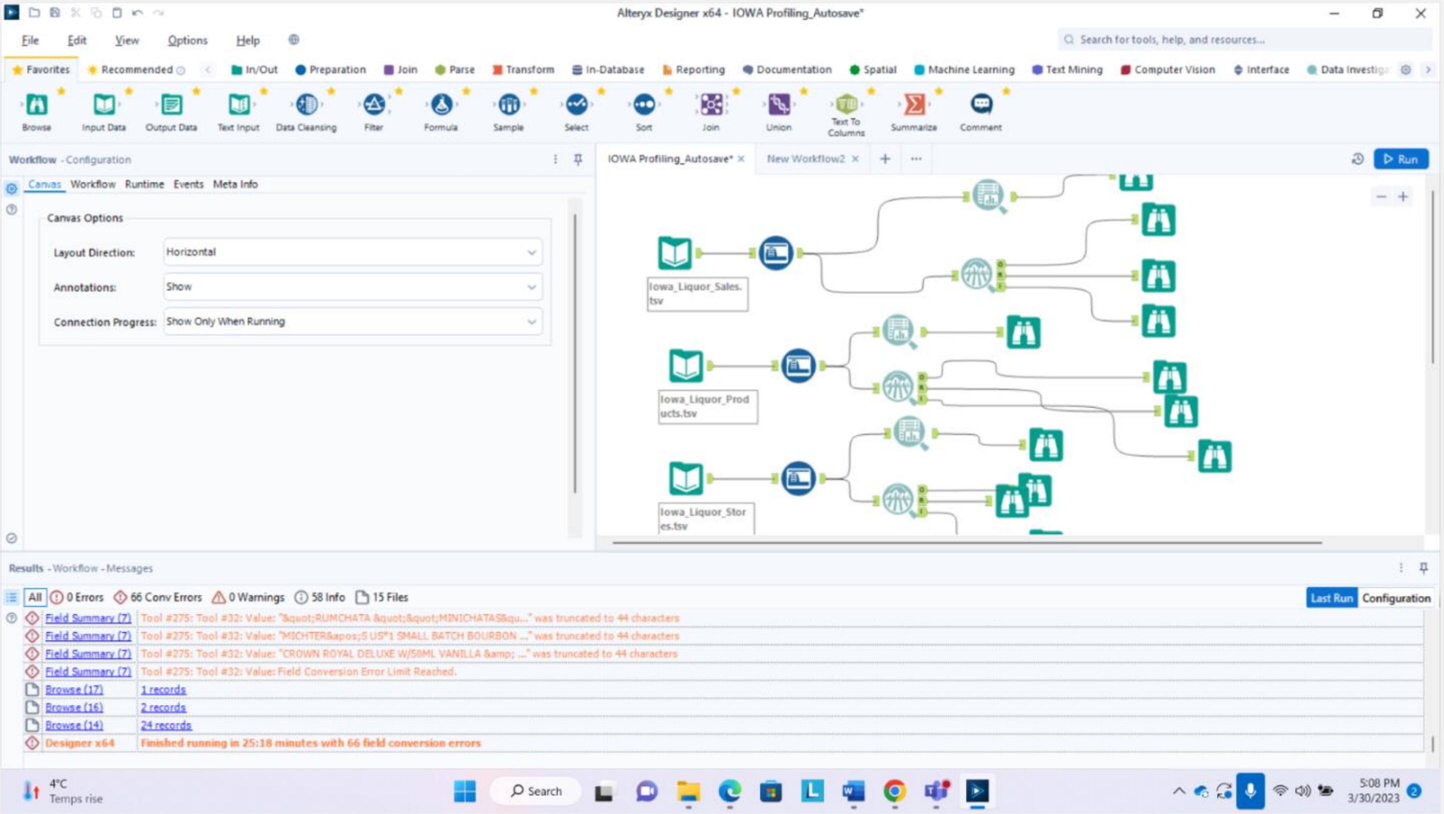
DATA PROFILING:



Canvas Options

Layout Direction: Horizontal

Annotations: Show

Connection Progress: Show Only When Running

All | 0 Errors | 66 Conv Errors | 0 Warnings | 58 Info | 15 Files

```
Field Summary (7)  Tool #275: Tool #32: Value: "&quot;RUMCHATA &quot;&quot;MINICHATAS&qu..." was truncated to 44 characters
Field Summary (7)  Tool #275: Tool #32: Value: "MICHTER&apos;S US*1 SMALL BATCH BOURBON ..." was truncated to 44 characters
Field Summary (7)  Tool #275: Tool #32: Value: "CROWN ROYAL DELUXE W/50ML VANILLA &amp; ..." was truncated to 44 characters
Field Summary (7)  Tool #275: Tool #32: Value: Field Conversion Error Limit Reached.
Browse (17)        1 records
Browse (16)        2 records
Browse (14)        24 records
Designer x64       Finished running in 25:18 minutes with 66 field conversion errors
```

We listed down column names and data types that were coming from staging tables so it was easier while loading into dim and facts

IOWA Sales Data

String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Item Number	0.0%	11,730	633	x904631	1	248,845	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.

County Number	2.9%	100	90	90	4,310	4,459,618	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Zip Code	0.3%	510	52501	52501	1	599,545	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Address	0.3%	2,822	123 A ST	1510 SOUTH ANKENY BOULEVARD PRAIRIE TRAIL SUITE 106 108	1	279,983	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Store Name	0.0%	2,940	SAME	FORT MADISON TOBACCO AND LIQUOR OUTLET PLUS / FORT MADISON	1	205,995	Some values of this field have a small number of value

							counts. If Appropriate, consider combining some value levels together.
County	0.6%	105	IDA	POTTAWATTAMIE	2	4,568,617	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Store Number	0.0%	2,807	5307	010025	1	205,995	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
City	0.3%	477	ELY	COLUMBUS JUNCTION	1	2,135,584	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.

Store Location	9.9%	3,204	POINT (-91.02618 41.59)	POINT (-91.67707300000001 41.968016000000006)	1	2,450,072	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Date	0.0%	2,730	12/22/2016	12/22/2016	1	18,002	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Vendor Number	0.0%	439	55	420	1	4,127,854	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Category Name	0.1%	114	MEZCAL	AMERICAN DISTILLED SPIRITS SPECIALTY	1	2,376,643	Some values of this field have a small number of value

							counts. If Appropriate, consider combining some value levels together.
Vendor Name	0.0%	549	MHW LTD	PRESTIGE WINE & SPIRITS GROUP / UNITED STATES DISTILLED PRODUCTS CO	1	4,127,854	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Item Description	0.0%	10,745	A+	THE GLENLIVET TASTING KIT W/12YR 200ML, 15YR 200ML & 18YR 200ML	1	631,415	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Invoice/Item Number	0.0%	24,847,481	378900006	RINV-04236500053	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.

IOWA Product

String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
List Date	0.0%	720	10/01/2017	10/01/2017	1	205	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Vendor	0.0%	280	421	421	1	728	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Category Name	0.0%	49	MEZCAL	IMPORTED DISTILLED SPIRITS SPECIALTY	3	644	Some values of this field have a small number of value counts. If Appropriate, consider

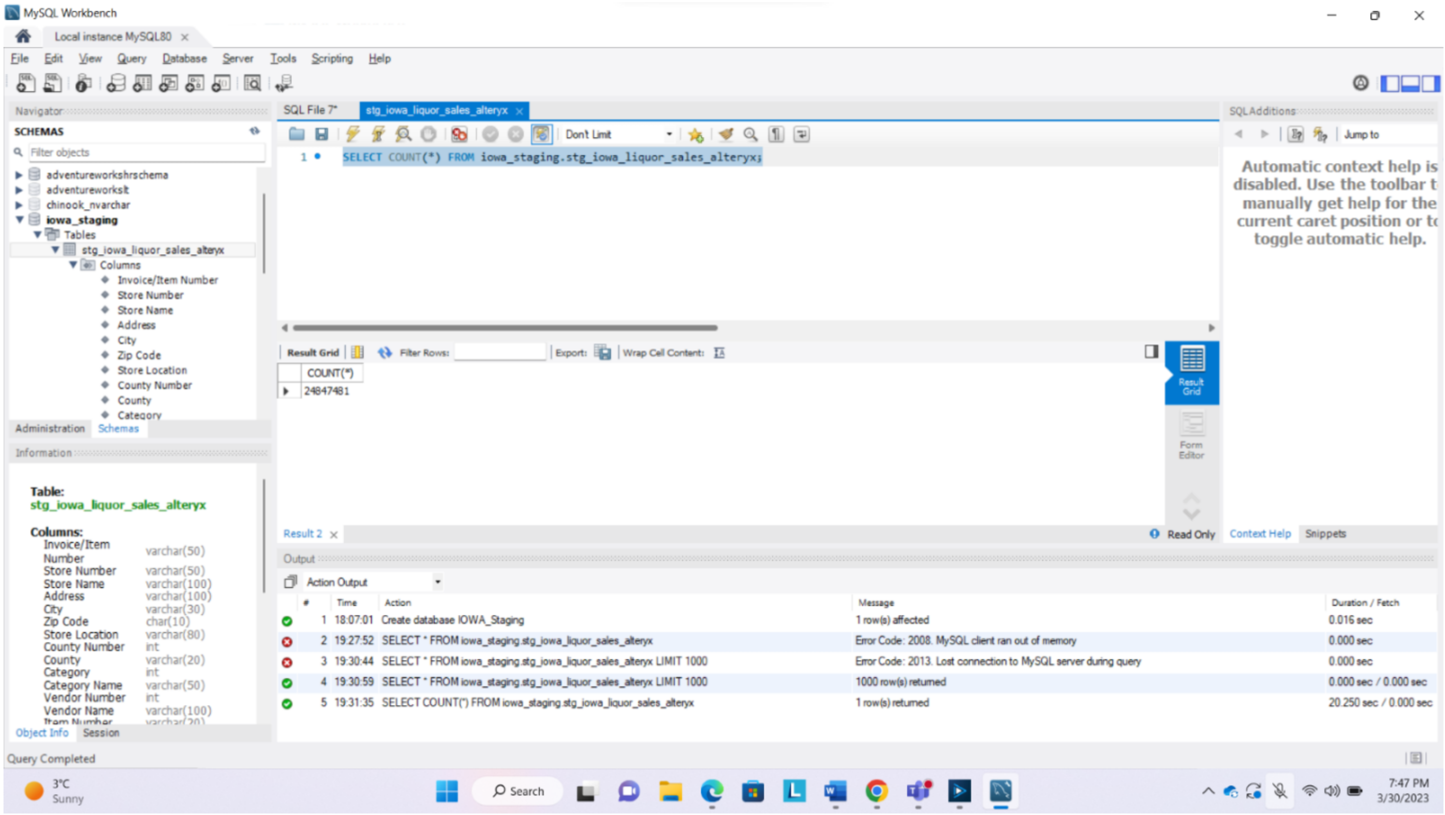
							combining some value levels together.
Vendor Name	0.0%	280	PROXIMO	AMERICAN HERITAGE DISTILLERS, LLC / CENTURY FARMS DISTILLERY	1	728	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Item Description	0.0%	5,243	P31	HIGH WEST WHISKEY HUCKBERRY GIFT PACK W/WHISKEY PEAK GLASSES	1	8	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Report Date	0.0%	1	10/01/2022	10/01/2022	6,350	6,350	

String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
Address	0.0%	2,198	708 MAIN	1510 SOUTH ANKENY BOULEVARD PRAIRIE TRAIL	1	3	Some values of this field have a small number of value

				SUITE 106 108			counts. If Appropriate, consider combining some value levels together.
Store Address	5.4%	1,950	POINT (-91.02618 41.59)	POINT (-93.755906 41.623602)	1	124	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Name	0.0%	2,237	RNDC	FORT MADISON TOBACCO AND LIQUOR OUTLET PLUS / FORT MADISON	1	3	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
City	0.0%	458	ELY	COLUMBUS JUNCTION	1	106	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.

Store	0.0%	2,277	5529	010035	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Store Status	0.0%	2	A	A	323	1,954	
State	0.0%	1	IA	IA	2,277	2,277	
Report Date	0.0%	1	10/04/2022	10/04/2022	2,277	2,277	



```
SELECT COUNT(*) FROM iowa_staging.stg_iowa_liquor_sales_alteryx;
```

COUNT(*)
24847481

Table:
stg_iowa_liquor_sales_alteryx

Columns:

Invoice/Item Number	varchar(50)
Store Number	varchar(50)
Store Name	varchar(100)
Address	varchar(100)
City	varchar(30)
Zip Code	char(10)
Store Location	varchar(80)
County Number	int
County	varchar(20)
Category	int
Category Name	varchar(50)
Vendor Number	int
Vendor Name	varchar(100)
Item Number	varchar(20)

```
#  Time      Action                                                               Message                                                  Duration / Fetch
1  18:07:01  Create database IOWA_Staging                                         1 row(s) affected                                        0.016 sec
2  19:27:52  SELECT * FROM iowa_staging.stg_iowa_liquor_sales_alteryx             Error Code: 2008. MySQL client ran out of memory         0.000 sec
3  19:30:44  SELECT * FROM iowa_staging.stg_iowa_liquor_sales_alteryx LIMIT 1000  Error Code: 2013. Lost connection to MySQL server during query  0.000 sec
4  19:30:59  SELECT * FROM iowa_staging.stg_iowa_liquor_sales_alteryx LIMIT 1000  1000 row(s) returned                                     0.000 sec / 0.000 sec
5  19:31:35  SELECT COUNT(*) FROM iowa_staging.stg_iowa_liquor_sales_alteryx      1 row(s) returned                                        20.250 sec / 0.000 sec
```

Some Insights after doing data profiling part:

1) We found there was a discrepancy in the county column between sales table and county population table. In one of the tables the word county was present whereas it was absent in the other table, so we had to make it in a standard format to perform joins.

2) We also found that there were case sensitive issues. The data coming from the sales stage table had store names in capital case whereas store stage table had mixed case format. So, while loading the fact table we had to put up a data cleaning tool to make data consistent
3) By analyzing data with this query:
select storename, date, sum(bottle sold), sum(sales dollars) from sales group by storename, date
order by store name asc, date desc

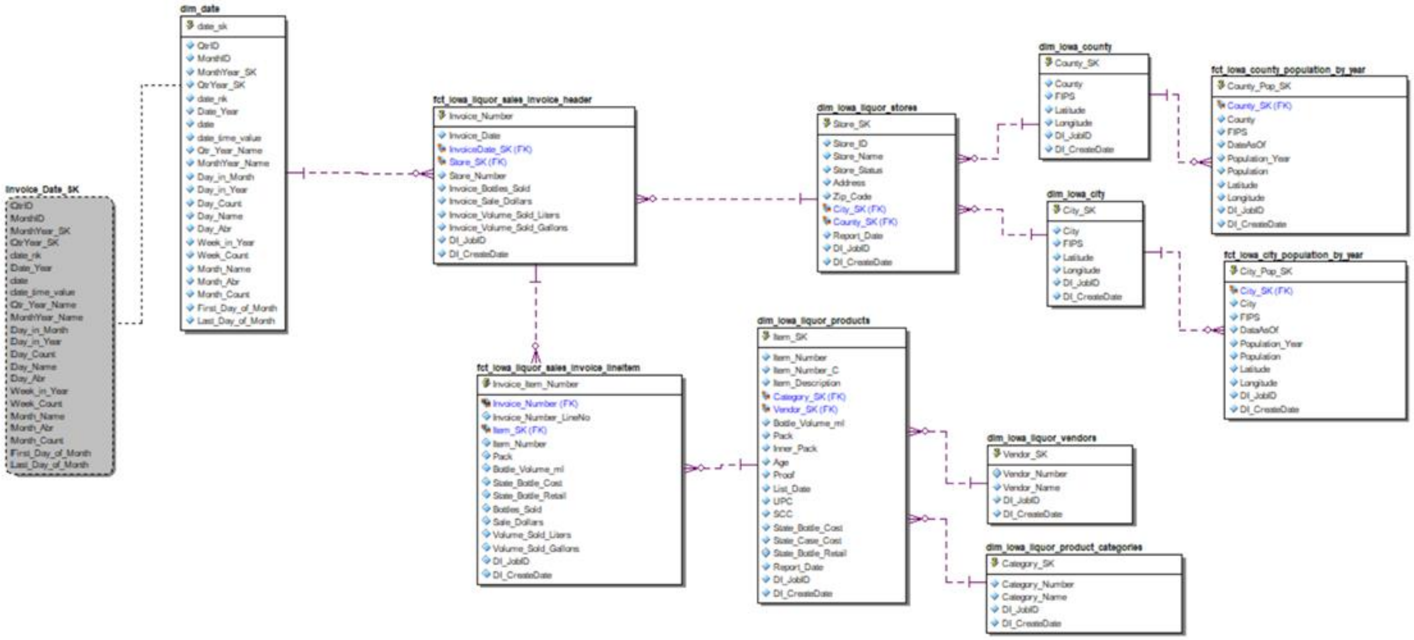
We found a pattern that for a every store name, there were many invoices generated on a particular date. Now using the summation tool we first populated header table. And then we populated line item fact table.

4) We also populated missing data with standard data like:

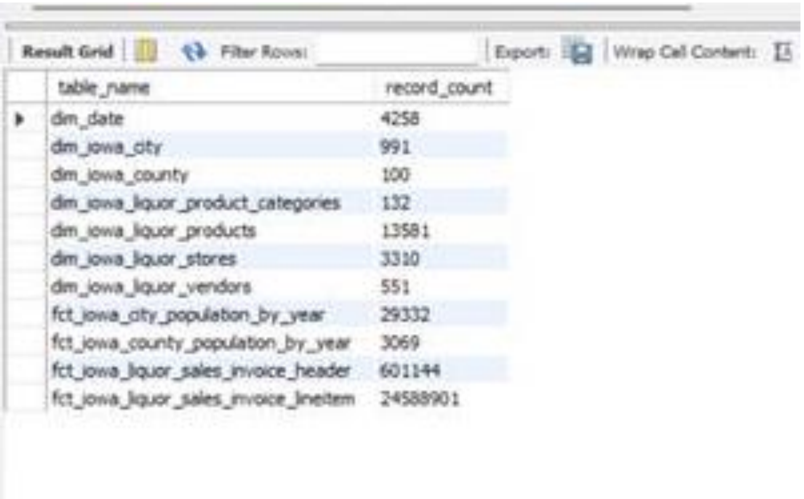
-99: null (integer data types)
'Unknown': null (varchar data types)
(0,0): location:null

5) The location column in all the staging tables had a value similar to POINT(longitude, latitude). For data viz we required (lat,long), so we had to split the data into 3 columns (location, latitude, longitude)

6) Reloaded dimension: For County, Category, Vendor, City data in the staging tables had missing values (that is few cities were present in sales staging tables which were missing in cities staging tables. So, to populate dimensions with accurate and all the cities, we performed union between the left and inner join)



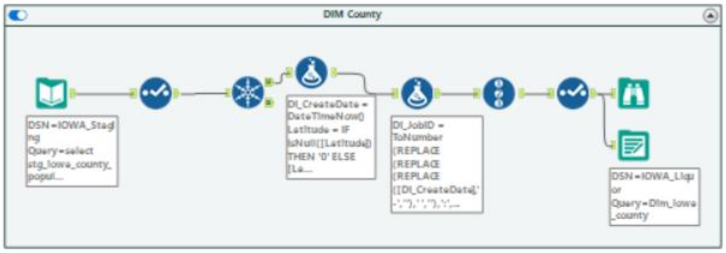
Row counts:



table_name	record_count
dim_date	4258
dim_iowa_city	991
dim_iowa_county	100
dim_iowa_liquor_product_categories	132
dim_iowa_liquor_products	13581
dim_iowa_liquor_stores	3310
dim_iowa_liquor_vendors	551
fct_iowa_city_population_by_year	29332
fct_iowa_county_population_by_year	3069
fct_iowa_liquor_sales_invoice_header	601144
fct_iowa_liquor_sales_invoice_lineitem	24588901

IOWA Part 2

Dim_County & Dim_Category_Reloaded:

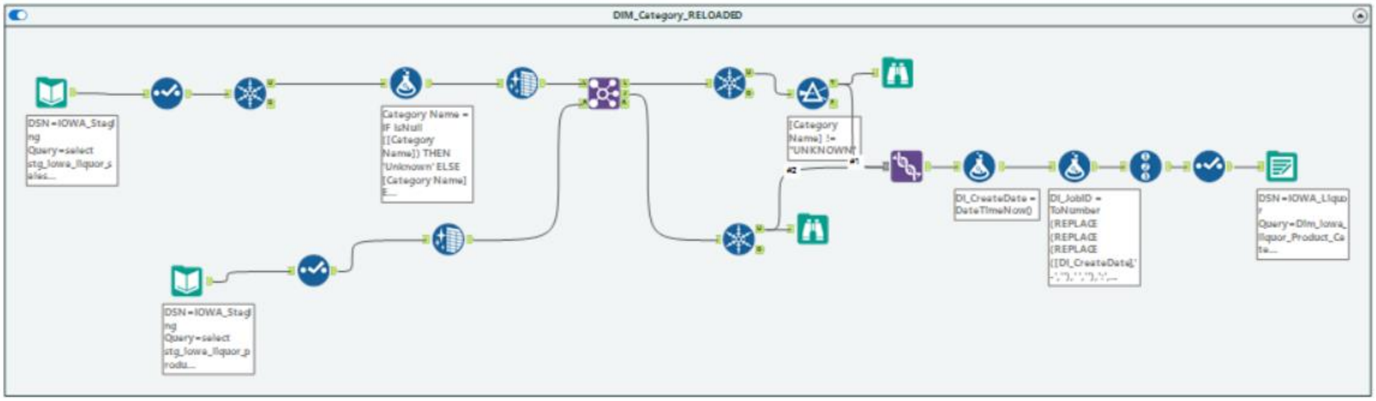


DSN=IOWA_Staging
Query=select stg_Iowa_county_popul...

DI_CreateDate = DateTimeNow()
Latitude = IF IsNull([Latitude]) THEN '0' ELSE [La...

DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],'-','')',' ','')',':',...

DSN=IOWA_Liquor
Query=Dim_Iowa_county



DSN=IOWA_Staging
Query=select stg_Iowa_liquor_sales...

Category Name = IF IsNull ([Category Name]) THEN 'Unknown' ELSE [Category Name] E...

[Category Name] != "UNKNOWN"

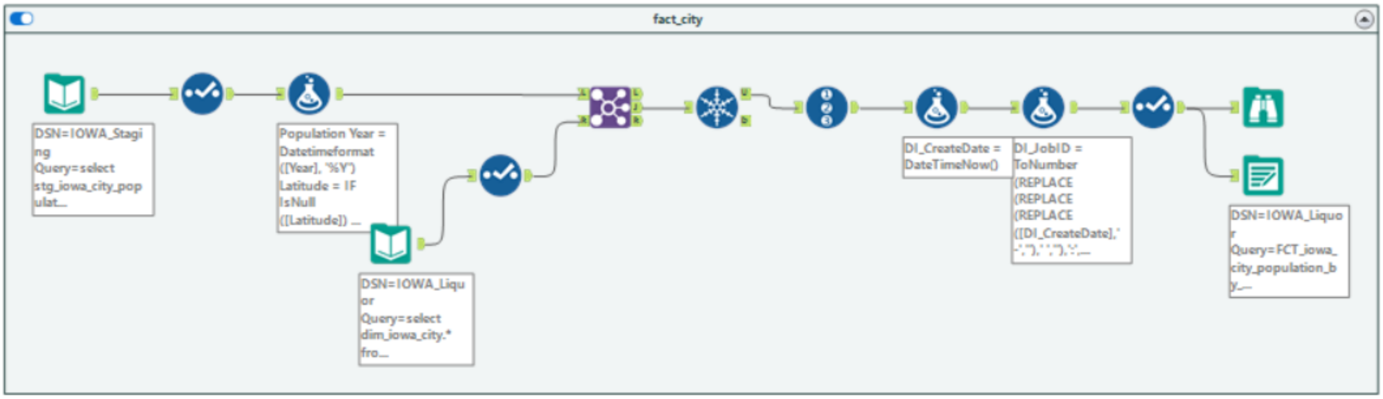
DI_CreateDate = DateTimeNow()

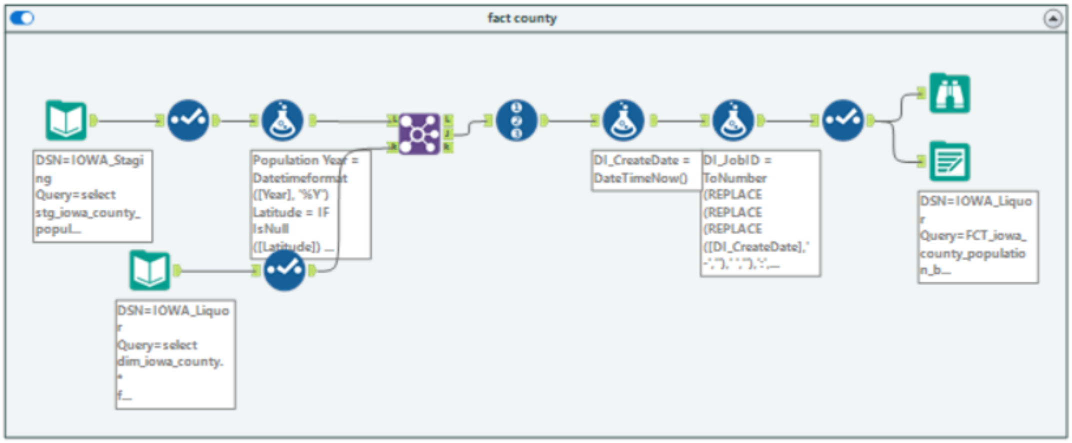
DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],'-','')',' ','')',':',...

DSN=IOWA_Liquor
Query=Dim_Iowa_liquor_Product_Cate...

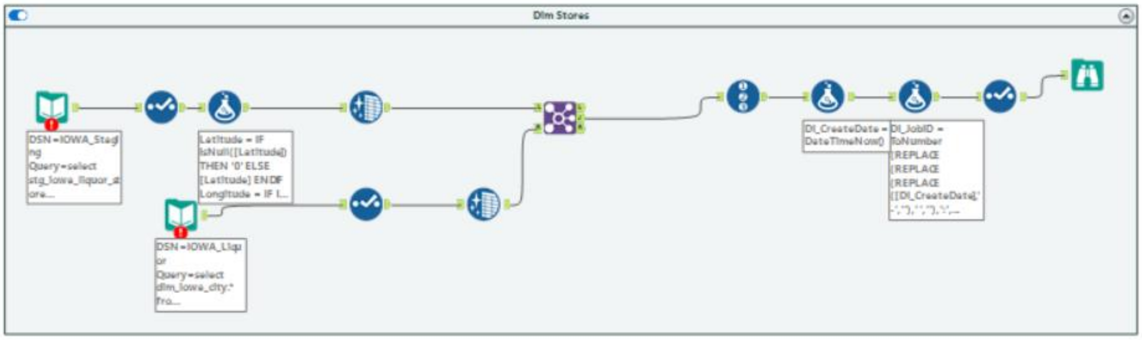
DSN=IOWA_Staging
Query=select stg_Iowa_liquor_produ...

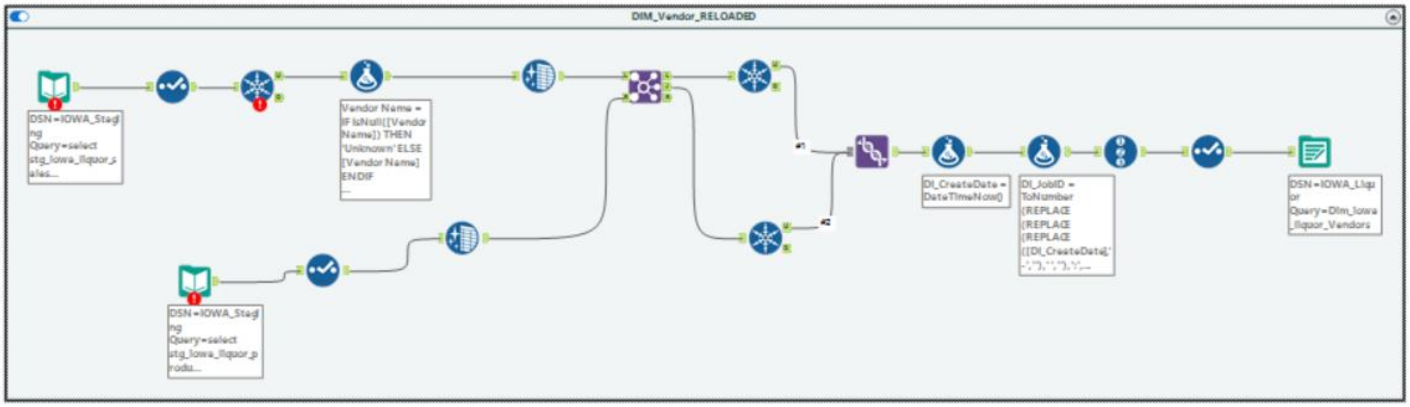
FCT_Total_Population_by_City & FCT_Total_Population_by_County:


fact_city
DSN=IOWA_Staging Query=select stg_iowa_city_populat...
Population Year = Datetimeformat ([Year], '%Y') Latitude = IF IsNull ([Latitude]) ...
DSN=IOWA_Liquor Query=select dim_iowa_city.* fro...
DI_CreateDate = DateTimeNow()
DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],' -',''),' ',''),':',...
DSN=IOWA_Liquor Query=FCT_iowa_city_population_by_...

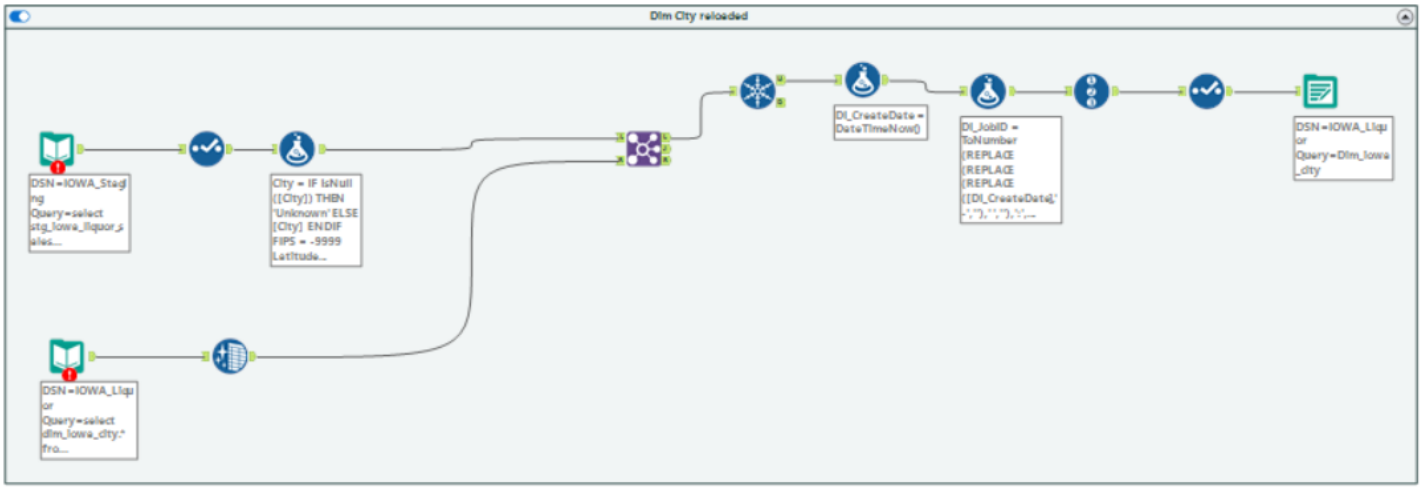

fact county
DSN=IOWA_Staging Query=select stg_iowa_county_popul...
Population Year = Datetimeformat ([Year], '%Y') Latitude = IF IsNull ([Latitude]) ...
DSN=IOWA_Liquor Query=select dim_iowa_county.* f...
DI_CreateDate = DateTimeNow()
DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],' -',''),' ',''),':',...
DSN=IOWA_Liquor Query=FCT_iowa_county_population_b...

Dim_Stores & Dim_Vendor_Reloaded:


Dim Stores
DSN=IOWA_Staging Query=select stg_iowa_liquor_store...
Latitude = IF IsNull([Latitude]) THEN '0' ELSE [Latitude] ENDIF Longitude = IF I...
DSN=IOWA_Liquor Query=select dim_iowa_city.* fro...
DI_CreateDate = DateTimeNow()
DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],' -',''),' ',''),':',...

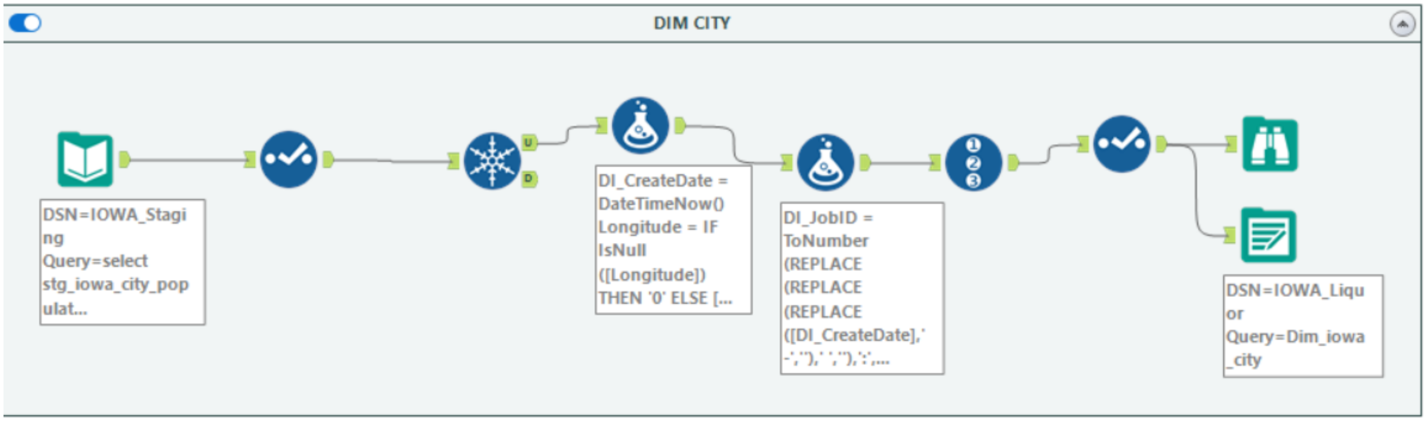

DIM_Vendor_RELOADED
DSN=IOWA_Staging Query=select stg_iowa_liquor_sales...
Vendor Name = IF IsNull([Vendor Name]) THEN 'Unknown' ELSE [Vendor Name] ENDIF ...
DSN=IOWA_Staging Query=select stg_iowa_liquor_produ...
#1
#2
DI_CreateDate = DateTimeNow()
DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],' -',''),' ',''),':',...
DSN=IOWA_Liquor Query=Dim_iowa_liquor_Vendors

Dim_City_Reloaded:



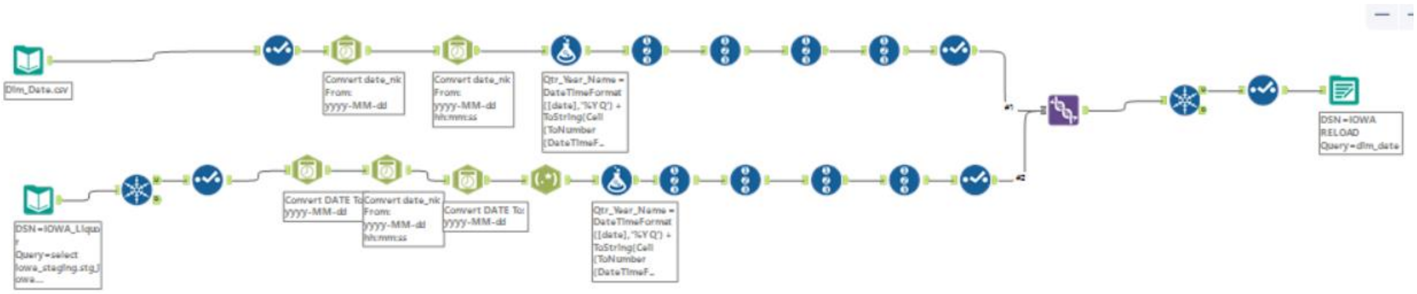
```
DSN=IOWA_Staging Query=select stg_iowa_liquor_sales...
City = IF IsNull ([City]) THEN 'Unknown' ELSE [City] ENDIF FIPS = -9999 Latitude...
DSN=IOWA_Liquor Query=select dim_iowa_city.* fro...
DI_CreateDate = DateTimeNow()
DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],'-',''),' ',''),':',...
DSN=IOWA_Liquor Query=Dim_iowa_city
```

Dim_City:



```
DSN=IOWA_Staging Query=select stg_iowa_city_populat...
DI_CreateDate = DateTimeNow() Longitude = IF IsNull ([Longitude]) THEN '0' ELSE [...
DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],'-',''),' ',''),':',...
DSN=IOWA_Liquor Query=Dim_iowa_city
```

Dim_Date:



Dim_Date.csv

Convert date_nk From: yyyy-MM-dd

Convert date_nk From: yyyy-MM-dd hh:mm:ss

Qtr_Year_Name = DateTimeFormat ([date],'%Y Q') + ToString(Ceil (ToNumber (DateTimeF...

DSN=IOWA_Liquor Query=select iowa_staging.stg_iowa...

Convert DATE To: yyyy-MM-dd

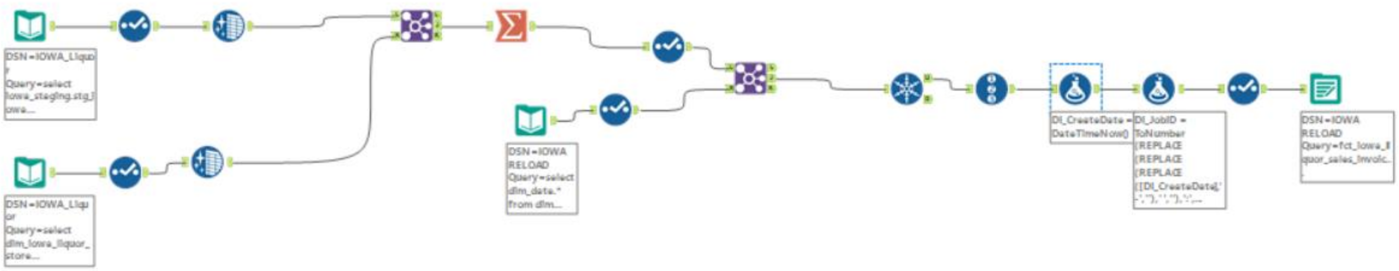
Convert date_nk From: yyyy-MM-dd hh:mm:ss

Convert DATE To: yyyy-MM-dd

Qtr_Year_Name = DateTimeFormat ([date],'%Y Q') + ToString(Ceil (ToNumber (DateTimeF...

DSN=IOWA RELOAD Query=dim_date

FCT_iowa_liquor_sales_header:



DSN=IOWA_Liquor Query=select iowa_staging.stg_iowa...

DSN=IOWA_Liquor Query=select dim_iowa_liquor_store...

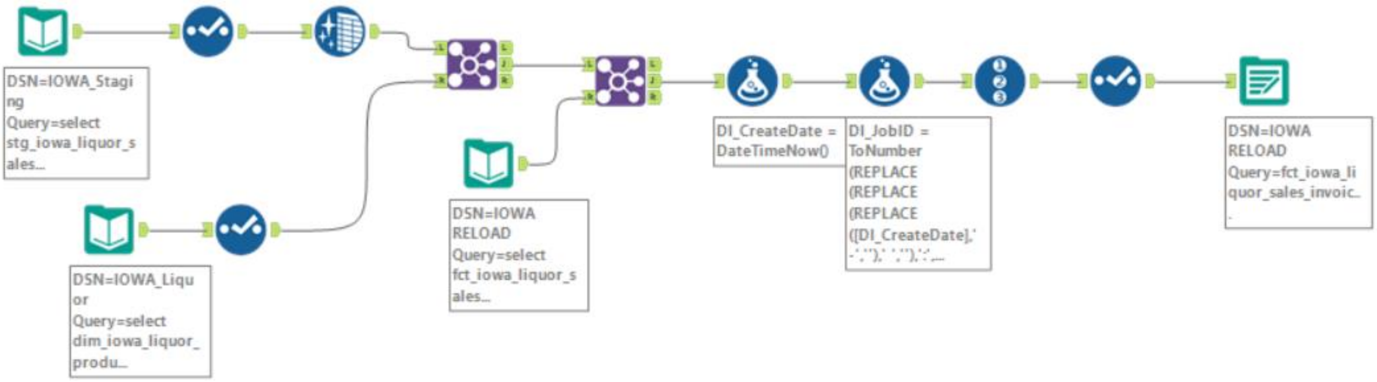
DSN=IOWA RELOAD Query=select dim_date.* from dim...

DI_CreateDate = DateTimeNow()

DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],' -','),' ',''),':',...

DSN=IOWA RELOAD Query=fct_iowa_liquor_sales_invoic...

FCT_iowa_liquor_sales_line_item:



DSN=IOWA_Staging Query=select stg_iowa_liquor_sales...

DSN=IOWA_Liquor Query=select dim_iowa_liquor_produ...

DSN=IOWA RELOAD Query=select fct_iowa_liquor_sales...

DI_CreateDate = DateTimeNow()

DI_JobID = ToNumber (REPLACE (REPLACE (REPLACE ([DI_CreateDate],' -',''),' ',''),':',...

DSN=IOWA RELOAD Query=fct_iowa_liquor_sales_invoic...