Shakir Mahabirsingh

Dimensional Lookup / Tableau KPI

Overview: This brief series of exercises were done to put into practice various skills learned about in Data Warehousing. Essentially, I transformed an OLTP database into a dimensional model (OLAP) by building an ETL system. I took a simple dataset containing Sales Representative info, built an ETL Transformation to load and manage my data, created a Dimensional Lookup/Update table, tested various changes, then loaded the final change into Tableau and visualized two KPIs. **These exercises were done by following Richard Holowczak's tutorials.**

1. This step was to create a Dimensional Lookup table after importing the [Salesrep] data. I first added this new transformation, imported the data, made sure each field is in the correct Type, Format, and Length. Then added the [Salesrep] dimension, connected the output to input, set the first table as the target table, and set the field source data and target dimension. Then I ran and reviewed the data in SQL Server.

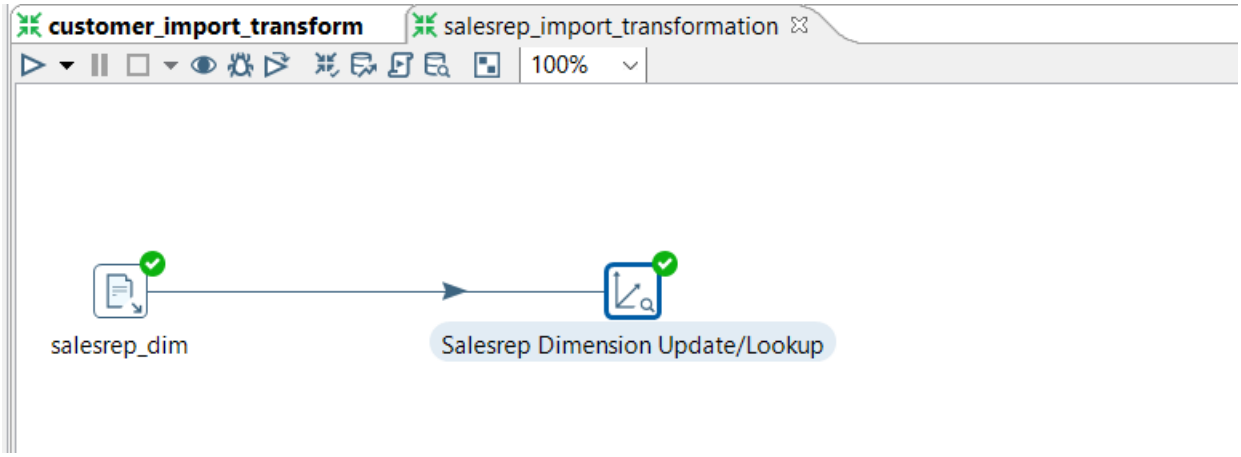
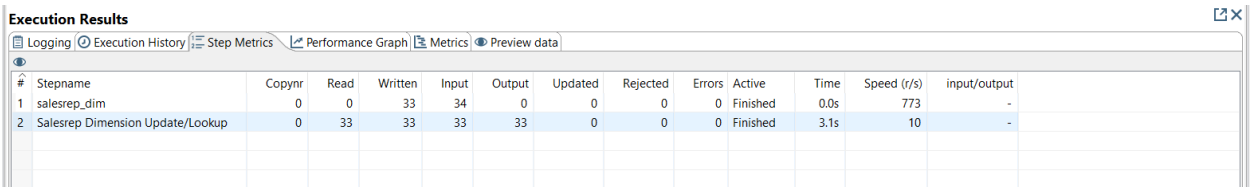


Figure 1 - Transformation Screen



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	salesrep_dim	0	0	33	34	0	0	0	0	Finished	0.0s	773	-
2	Salesrep Dimension Update/Lookup	0	33	33	33	33	0	0	0	Finished	3.1s	10	-

Figure 2 - Step Metrics

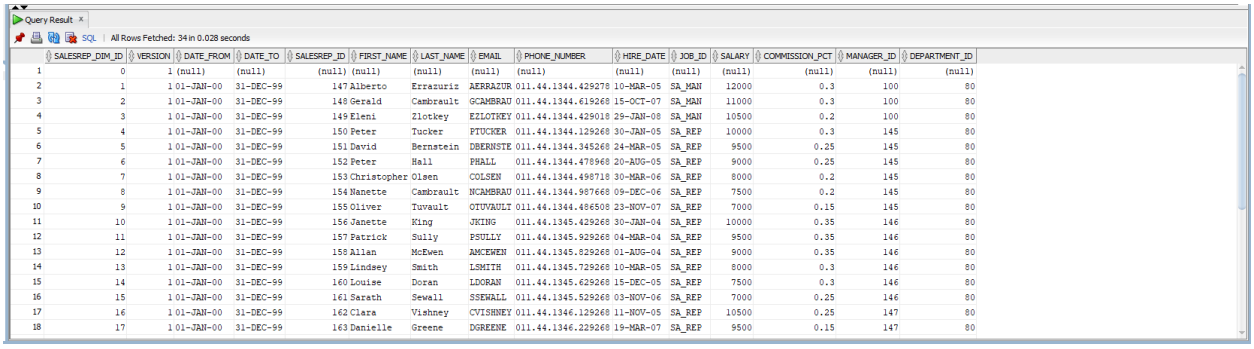


Figure 3 - Resulting Data

2. After importing the data file and establishing the proper Dimension Table, this step was to implement a Type 1 SCD. I performed the same steps as before but changed the Type of Dimension Update to “Punch through”. This allows a Type 1 SCD on these fields.

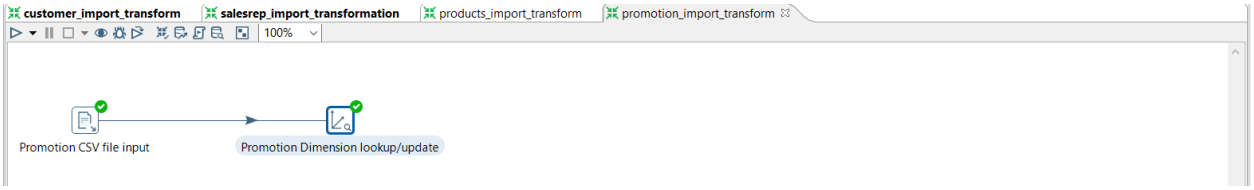


Figure 4 – Transformation Screen

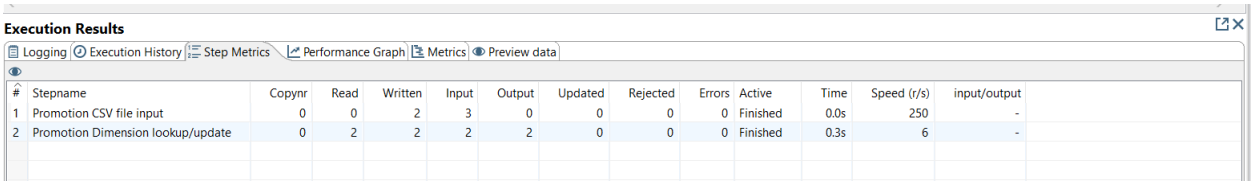


Figure 5 - Step Metrics

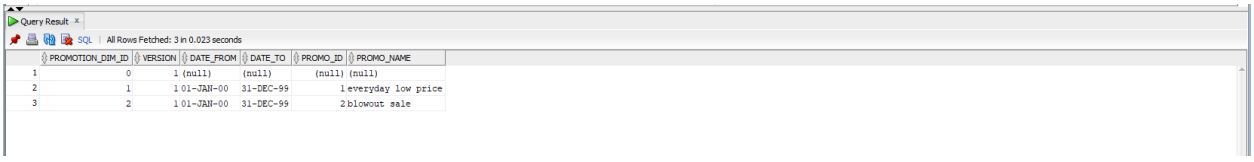


Figure 6 - Resulting Data

3. For this step I had to embellish a created Date Dimension with additional date attributes. To generate the Date Dimension it required first to create a table with the Start Date, link to a table that incrementally generated dates, link to an additional table that calculated the dates, then to a table that selected the values. After creating the various tables, I created one more table to calculate the additional fields and linked that to the final Dimensional Lookup.

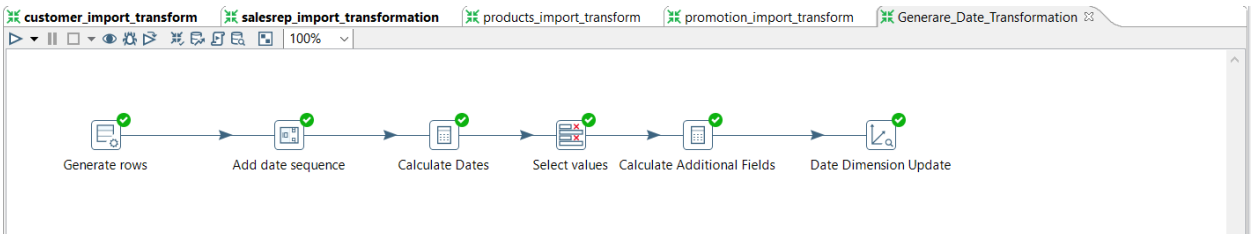


Figure 7 – Transformation Screen

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Generate rows	0	0	1100	0	0	0	0	0	Finished	0.0s	73,333	-
2	Add date sequence	0	1100	1100	0	0	0	0	0	Finished	0.0s	45,833	-
3	Calculate Dates	0	1100	1100	0	0	0	0	0	Finished	0.1s	12,941	-
4	Select values	0	1100	1100	0	0	0	0	0	Finished	0.2s	7,285	-
5	Calculate Additional Fields	0	1100	1100	0	0	0	0	0	Finished	0.2s	5,392	-
6	Date Dimension Update	0	1100	1100	1100	1100	0	0	0	Finished	1mn 3s	17	-

Figure 8 - Step Metrics

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

#	Sales_Date	sales_day_of_year	sales_month	sales_year	sales_quarter	sales_month_name	sales_day_of_week_name	sales_day_of_week	sales_day_of_month	date_dim_id
1	01/02/2006	2	1	2006	1	January	Monday	2	2	1101
2	01/03/2006	3	1	2006	1	January	Tuesday	3	3	1102
3	01/04/2006	4	1	2006	1	January	Wednesday	4	4	1103
4	01/05/2006	5	1	2006	1	January	Thursday	5	5	1104
5	01/06/2006	6	1	2006	1	January	Friday	6	6	1105
6	01/07/2006	7	1	2006	1	January	Saturday	7	7	1106
7	01/08/2006	8	1	2006	1	January	Sunday	1	8	1107
8	01/09/2006	9	1	2006	1	January	Monday	2	9	1108
9	01/10/2006	10	1	2006	1	January	Tuesday	3	10	1109
1	01/11/2006	11	1	2006	1	January	Wednesday	4	11	1110

Figure 9 - Resulting Data

4. Here I had to embellish the Date Dimension with additional Holiday info. I loaded this data into a table, connected it separately to a created Stream Lookup table to actually check the various dates, then implemented these into the screen.

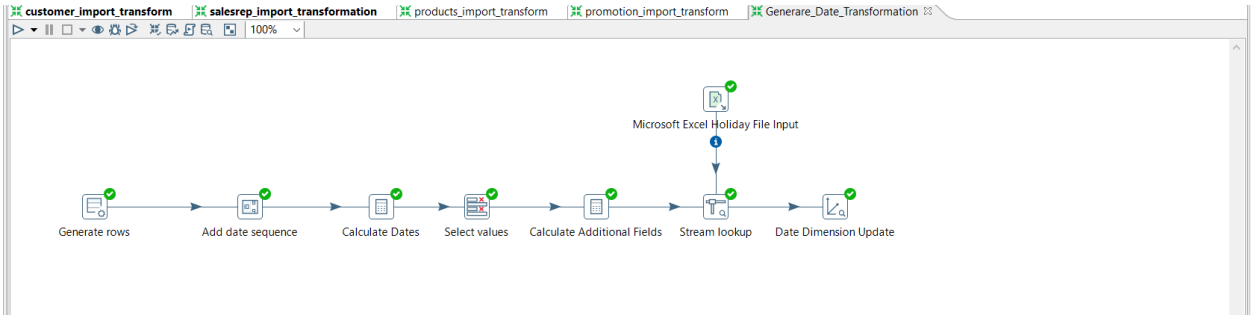
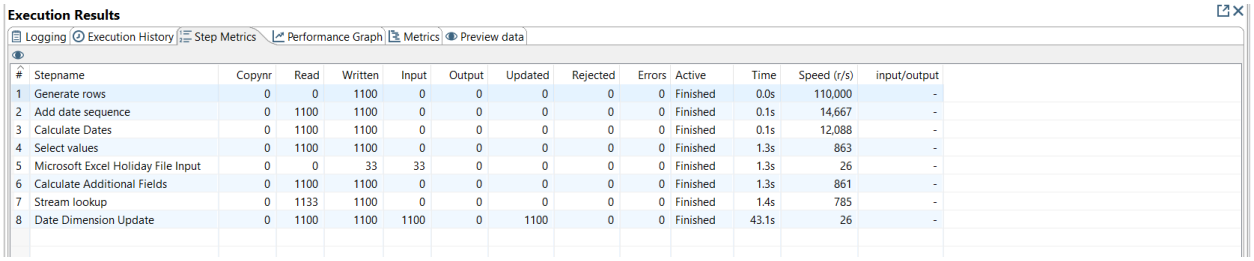


Figure 10 - Transaction Screen

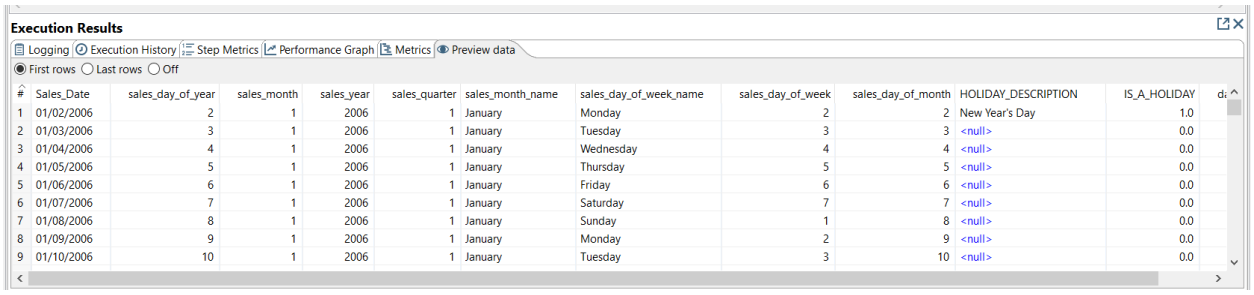


Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Generate rows	0	0	1100	0	0	0	0	0	Finished	0.0s	110,000	-
2	Add date sequence	0	1100	1100	0	0	0	0	0	Finished	0.1s	14,667	-
3	Calculate Dates	0	1100	1100	0	0	0	0	0	Finished	0.1s	12,088	-
4	Select values	0	1100	1100	0	0	0	0	0	Finished	1.3s	863	-
5	Microsoft Excel Holiday File Input	0	0	33	33	0	0	0	0	Finished	1.3s	26	-
6	Calculate Additional Fields	0	1100	1100	0	0	0	0	0	Finished	1.3s	861	-
7	Stream lookup	0	1133	1100	0	0	0	0	0	Finished	1.4s	785	-
8	Date Dimension Update	0	1100	1100	1100	0	1100	0	0	Finished	43.1s	26	-

Figure 11 - Step Metrics



Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

#	Sales_Date	sales_day_of_year	sales_month	sales_year	sales_quarter	sales_month_name	sales_day_of_week_name	sales_day_of_week	sales_day_of_month	HOLIDAY_DESCRIPTION	IS_A_HOLIDAY	di
1	01/02/2006	2	1	2006	1	January	Monday	2	2	New Year's Day	1.0	
2	01/03/2006	3	1	2006	1	January	Tuesday	3	3	<null>	0.0	
3	01/04/2006	4	1	2006	1	January	Wednesday	4	4	<null>	0.0	
4	01/05/2006	5	1	2006	1	January	Thursday	5	5	<null>	0.0	
5	01/06/2006	6	1	2006	1	January	Friday	6	6	<null>	0.0	
6	01/07/2006	7	1	2006	1	January	Saturday	7	7	<null>	0.0	
7	01/08/2006	8	1	2006	1	January	Sunday	1	8	<null>	0.0	
8	01/09/2006	9	1	2006	1	January	Monday	2	9	<null>	0.0	
9	01/10/2006	10	1	2006	1	January	Tuesday	3	10	<null>	0.0	

Figure 12 - Resulting Data

5. This step required updating the set of transformations with new data. This happening after I fit the various transformations and dimensions together and creating a central fact table. I uploaded the data since I already set the transformations properly, it didn't require additional changes.

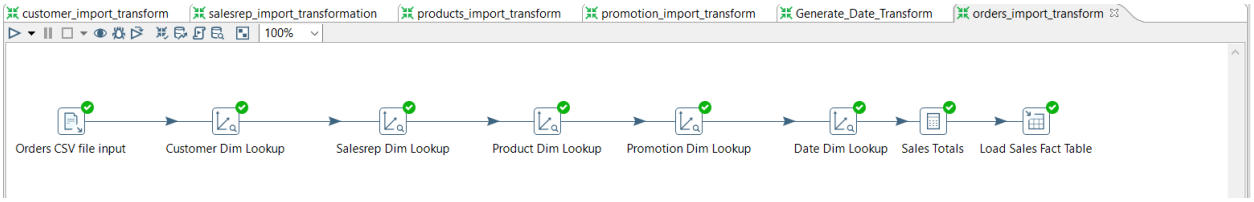
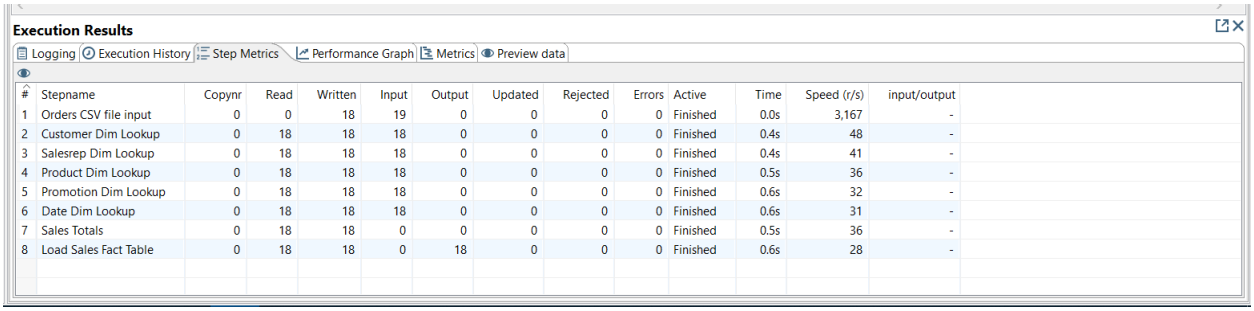


Figure 13 - Transaction Screen

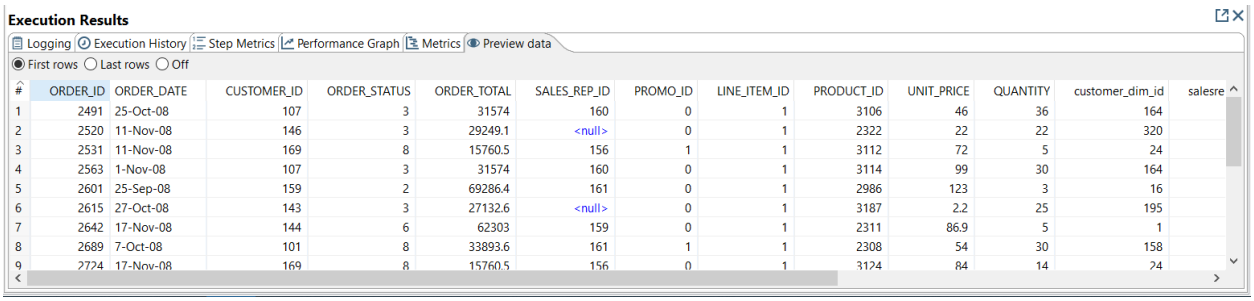


Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Orders CSV file input	0	0	18	19	0	0	0	0	Finished	0.0s	3,167	-
2	Customer Dim Lookup	0	18	18	18	0	0	0	0	Finished	0.4s	48	-
3	Salesrep Dim Lookup	0	18	18	18	0	0	0	0	Finished	0.4s	41	-
4	Product Dim Lookup	0	18	18	18	0	0	0	0	Finished	0.5s	36	-
5	Promotion Dim Lookup	0	18	18	18	0	0	0	0	Finished	0.6s	32	-
6	Date Dim Lookup	0	18	18	18	0	0	0	0	Finished	0.6s	31	-
7	Sales Totals	0	18	18	0	0	0	0	0	Finished	0.5s	36	-
8	Load Sales Fact Table	0	18	18	0	18	0	0	0	Finished	0.6s	28	-

Figure 14 - Step Metrics



Execution Results

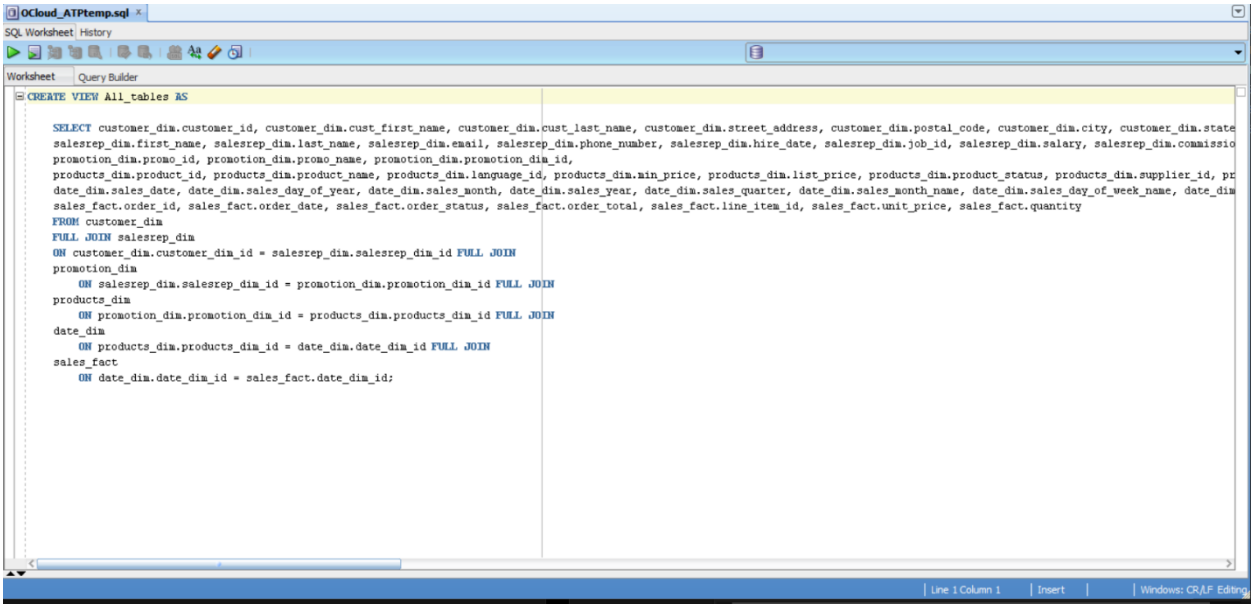
Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

#	ORDER_ID	ORDER_DATE	CUSTOMER_ID	ORDER_STATUS	ORDER_TOTAL	SALES_REP_ID	PROMO_ID	LINE_ITEM_ID	PRODUCT_ID	UNIT_PRICE	QUANTITY	customer_dim_id	salesre
1	2491	25-Oct-08	107	3	31574	160	0	1	3106	46	36	164	
2	2520	11-Nov-08	146	3	29249.1	<null>	0	1	2322	22	22	320	
3	2531	11-Nov-08	169	8	15760.5	156	1	1	3112	72	5	24	
4	2563	1-Nov-08	107	3	31574	160	0	1	3114	99	30	164	
5	2601	25-Sep-08	159	2	69286.4	161	0	1	2986	123	3	16	
6	2615	27-Oct-08	143	3	27132.6	<null>	0	1	3187	2.2	25	195	
7	2642	17-Nov-08	144	6	62303	159	0	1	2311	86.9	5	1	
8	2689	7-Oct-08	101	8	33893.6	161	1	1	2308	54	30	158	
9	2724	17-Nov-08	169	8	15760.5	156	0	1	3124	84	14	24	

Figure 15 - Resulting Data

SQL For Joining and Creating View



6. Finally, I loaded my data into Tableau and tested various visualizations for my below KPIs.

KPI Visualization

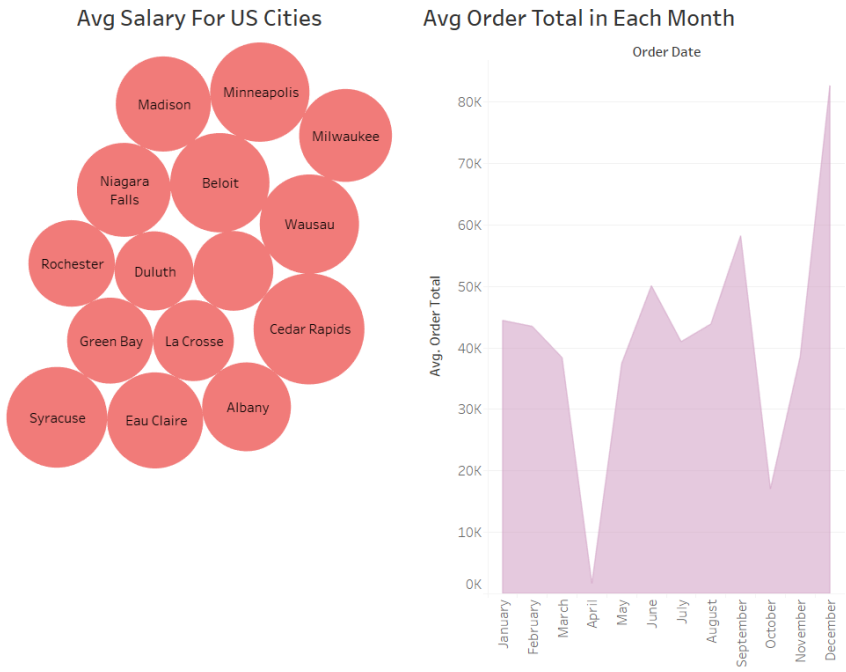


*Figure 16 – KPIs *Blank circle – Poughkeepsie (couldn't fit in screen)*

Conclusion:

- Recently I started tracking my time using an app (Toggl) and decided it would be perfect to track every minute I worked on this set of exercises. What I didn't realize, until reviewing the data when I finally finished, was that this took way longer than I could've imagined. From opening the tutorial and starting, it took about 6hrs. This isn't including the Tableau portion which was about an additional 1hr. The most difficult part was painstakingly making sure every field, table, connection, target table, everything was perfect. Numerous errors can occur and stop your Transaction Run if something (no matter how small) doesn't sync up. The last step of creating a view took me way longer than I should admit because I kept encountering a "duplicate column error". After searching far and wide, I finally realized the duplication was occurring in my Sales_Fact_Table. This single issue sums up what's difficult about the ETL and Database Process. Finally seeing all green checkmarks, data displaying correctly, and proper KPI visualizations flushes someone with a sense of relief like none other.