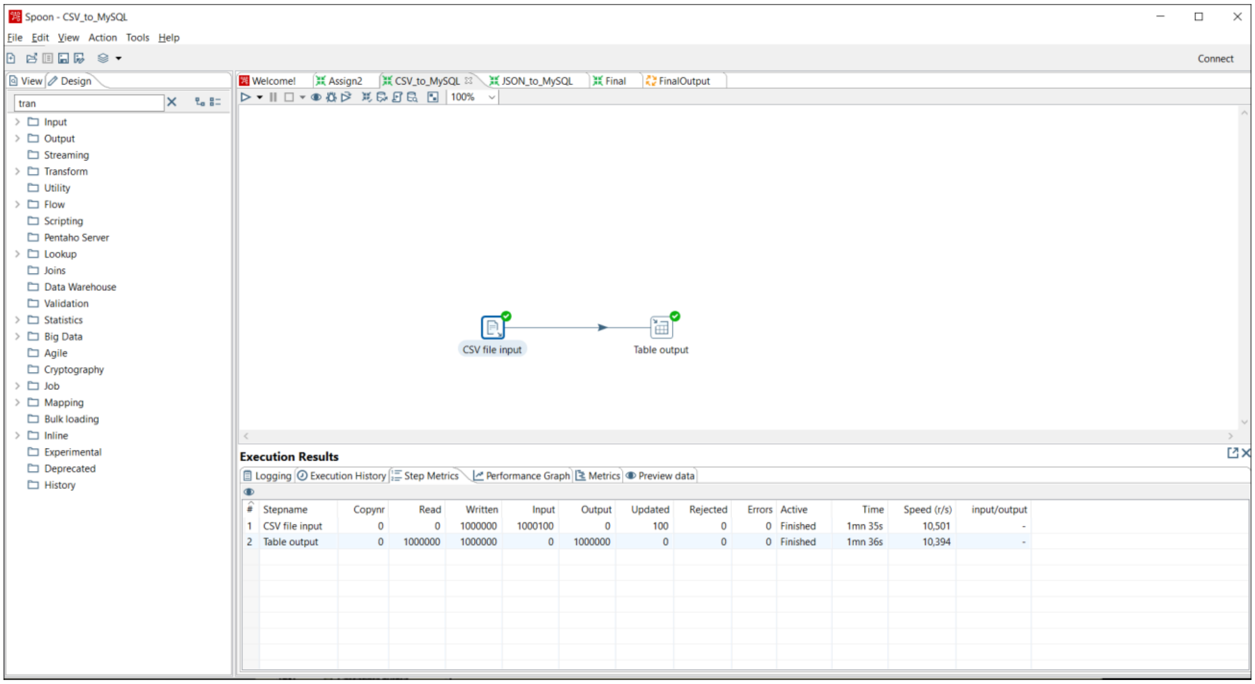
Madhu Jagdale Assignment 2

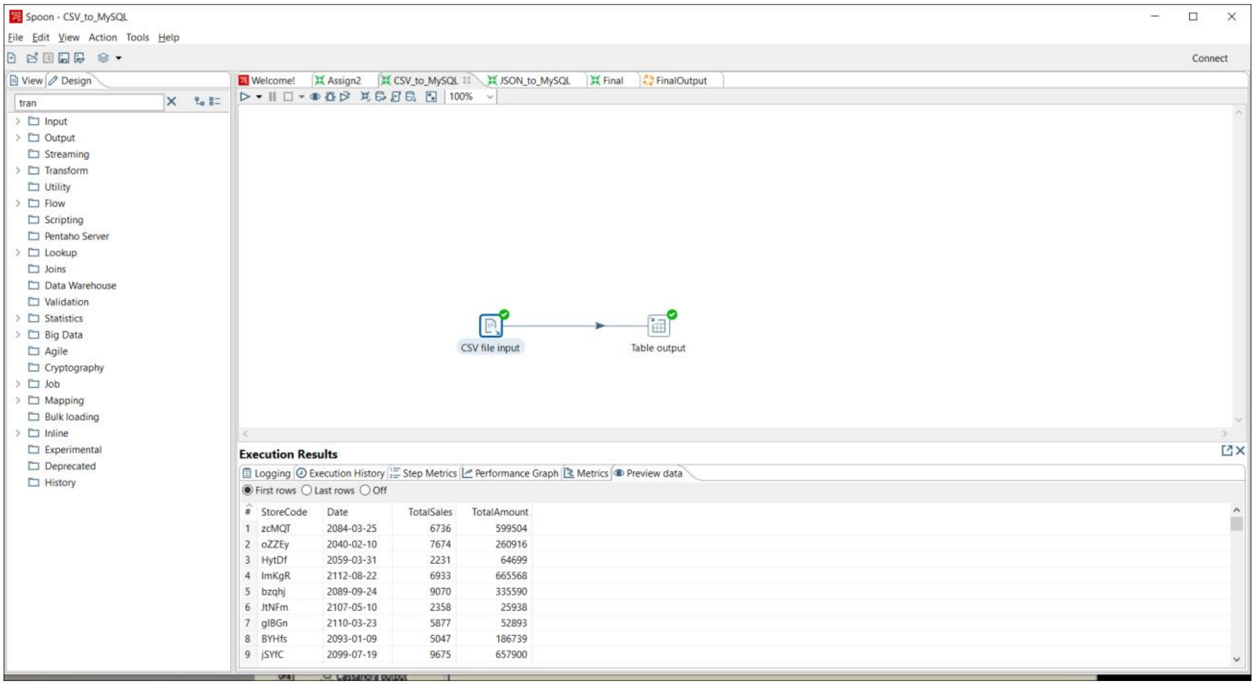
Perform Extraction:

2.a) Extract all the sales CSV files to a single MySQL table.



Execution Results

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	1000000	1000100	0	100	0	0	Finished	1mn 35s	10,501	-
2	Table output	0	1000000	1000000	0	1000000	0	0	0	Finished	1mn 36s	10,394	-

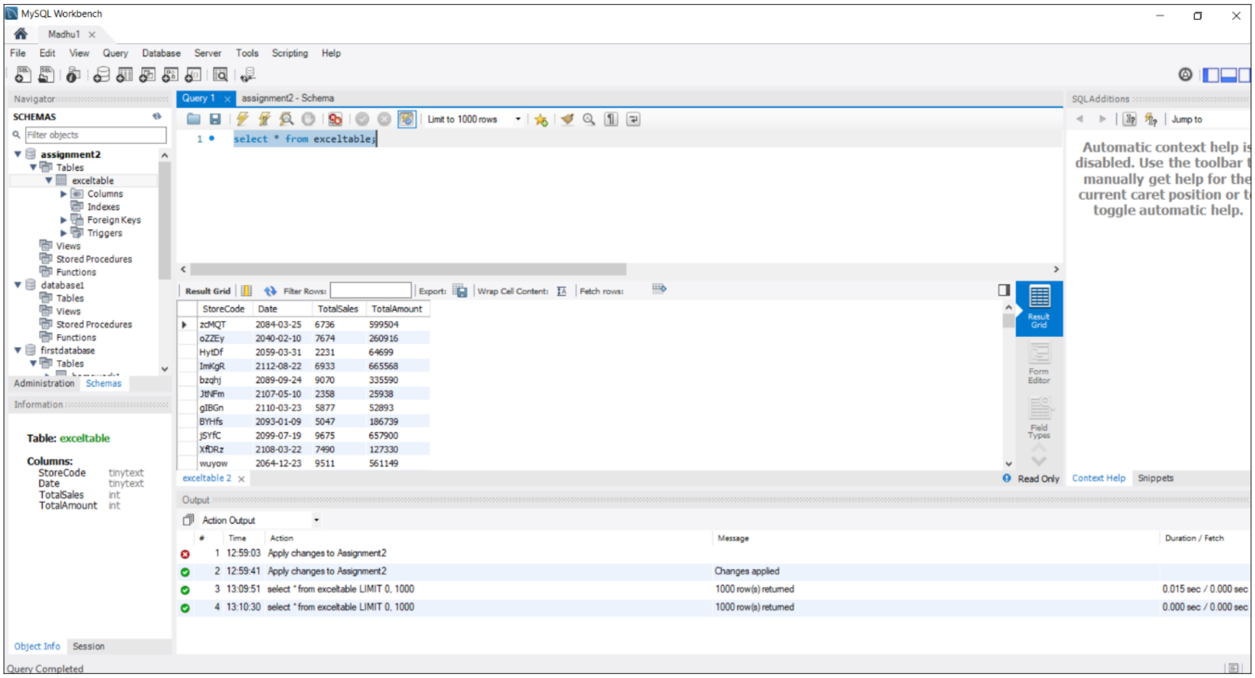


Execution Results

● First rows ○ Last rows ○ Off

#	StoreCode	Date	TotalSales	TotalAmount
1	zcMQT	2084-03-25	6736	599504
2	oZZEy	2040-02-10	7674	260916
3	HytDf	2059-03-31	2231	64699
4	ImKgR	2112-08-22	6933	665568
5	bzqhj	2089-09-24	9070	335590
6	JtNFm	2107-05-10	2358	25938
7	gIBGn	2110-03-23	5877	52893
8	BYHfs	2093-01-09	5047	186739
9	jSYfC	2099-07-19	9675	657900

Madhu Jagdale   Assignment 2



```
select * from exceltable;
```

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

Information

Table: exceltable

Columns:

StoreCode	tinytext
Date	tinytext
TotalSales	int
TotalAmount	int

StoreCode	Date	TotalSales	TotalAmount
zcMQT	2084-03-25	6736	599504
oZZEy	2040-02-10	7674	260916
HytDf	2059-03-31	2231	64699
ImKgR	2112-08-22	6933	665568
bzqhj	2089-09-24	9070	335590
JtNFm	2107-05-10	2358	25938
gIBGn	2110-03-23	5877	52893
BYHfs	2093-01-09	5047	186739
jSYfC	2099-07-19	9675	657900
XfDRz	2108-03-22	7490	127330
wuyow	2064-12-23	9511	561149

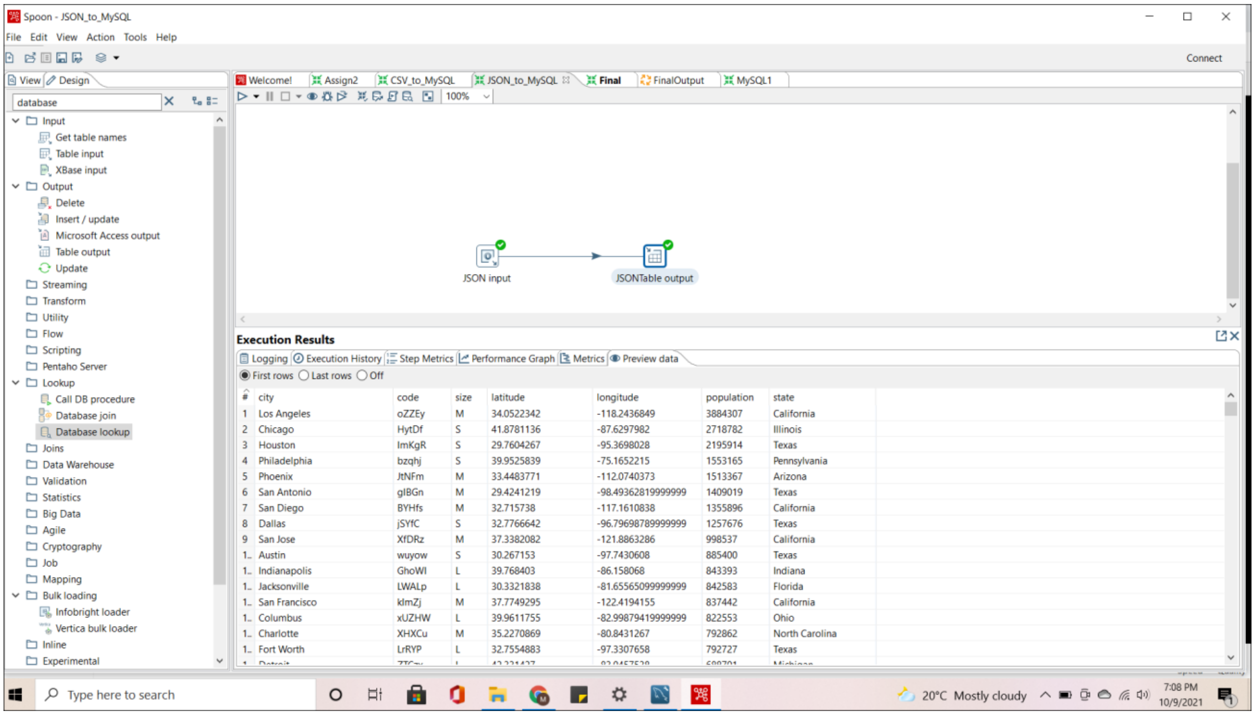
exceltable 2

Output

Action Output

#	Time	Action	Message	Duration / Fetch
1	12:59:03	Apply changes to Assignment2		
2	12:59:41	Apply changes to Assignment2	Changes applied	
3	13:09:51	select * from exceltable LIMIT 0, 1000	1000 row(s) returned	0.015 sec / 0.000 sec
4	13:10:30	select * from exceltable LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.000 sec

2.b)   Extract the store JSON file to another MySQL table



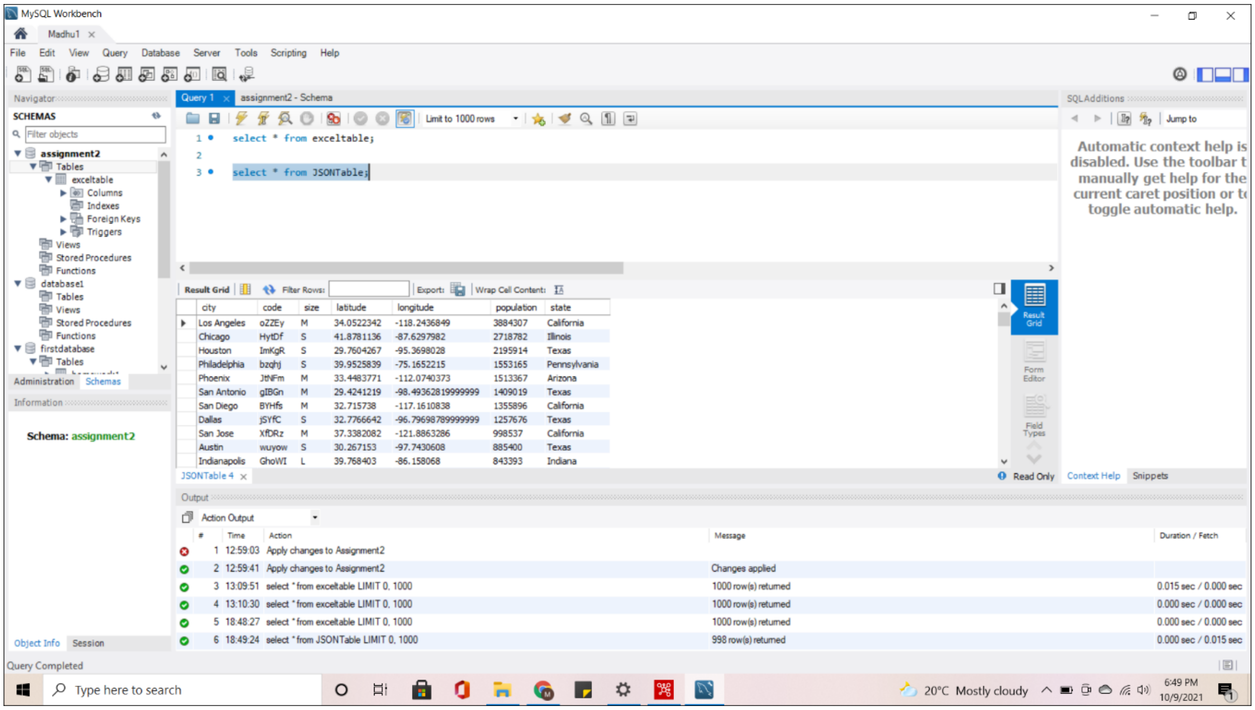
Execution Results

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

● First rows ○ Last rows ○ Off

#	city	code	size	latitude	longitude	population	state
1	Los Angeles	oZZEy	M	34.0522342	-118.2436849	3884307	California
2	Chicago	HytDf	S	41.8781136	-87.6297982	2718782	Illinois
3	Houston	ImKgR	S	29.7604267	-95.3698028	2195914	Texas
4	Philadelphia	bzqhj	S	39.9525839	-75.1652215	1553165	Pennsylvania
5	Phoenix	JtNFm	M	33.4483771	-112.0740373	1513367	Arizona
6	San Antonio	gIBGn	M	29.4241219	-98.49362819999999	1409019	Texas
7	San Diego	BYHfs	M	32.715738	-117.1610838	1355896	California
8	Dallas	jSYfC	S	32.7766642	-96.79698789999999	1257676	Texas
9	San Jose	XfDRz	M	37.3382082	-121.8863286	998537	California
1..	Austin	wuyow	S	30.267153	-97.7430608	885400	Texas
1..	Indianapolis	GhoWl	L	39.768403	-86.158068	843393	Indiana
1..	Jacksonville	LWALp	L	30.3321838	-81.65565099999999	842583	Florida
1..	San Francisco	klmZj	M	37.7749295	-122.4194155	837442	California
1..	Columbus	xUZHW	L	39.9611755	-82.99879419999999	822553	Ohio
1..	Charlotte	XHXCu	M	35.2270869	-80.8431267	792862	North Carolina
1..	Fort Worth	LrRYP	L	32.7554883	-97.3307658	792727	Texas

Madhu Jagdale Assignment 2



```
select * from exceltable;

select * from JSONTable;
```

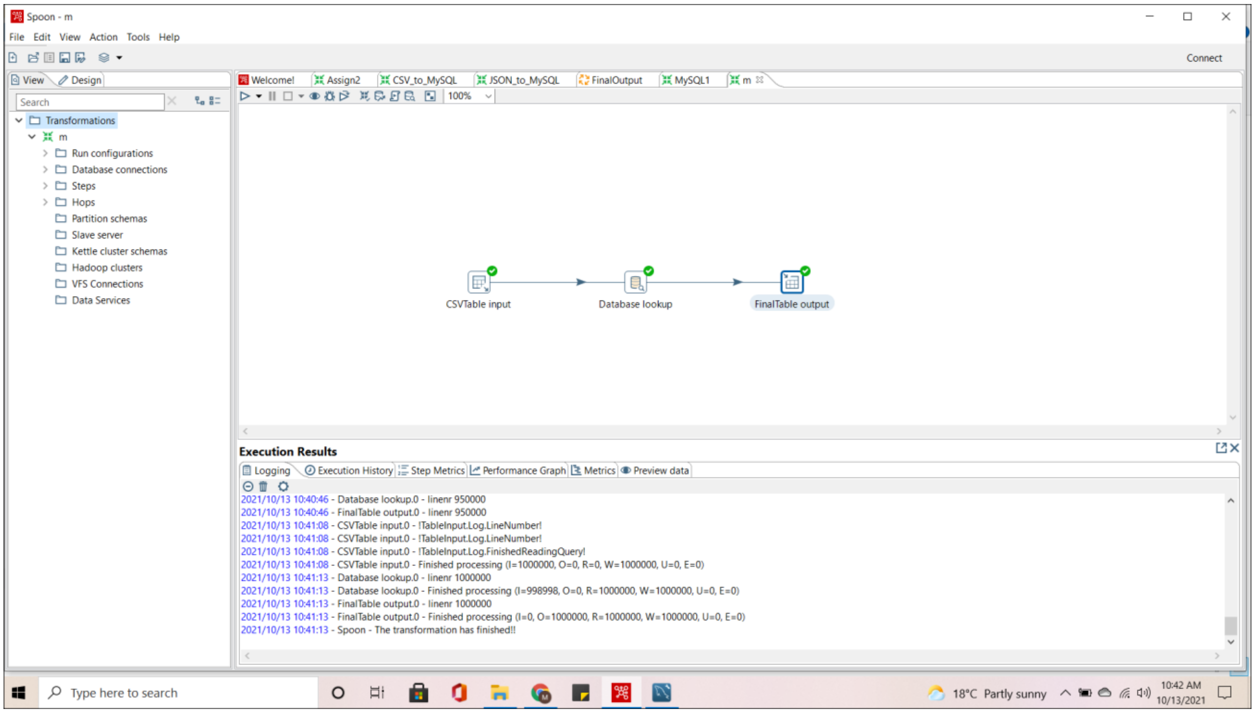
city	code	size	latitude	longitude	population	state
Los Angeles	oZZEy	M	34.0522342	-118.2436849	3884307	California
Chicago	HytDf	S	41.8781136	-87.6297982	2718782	Illinois
Houston	ImKgR	S	29.7604267	-95.3698028	2195914	Texas
Philadelphia	bzqhj	S	39.9525839	-75.1652215	1553165	Pennsylvania
Phoenix	JtNFm	M	33.4483771	-112.0740373	1513367	Arizona
San Antonio	gIBGn	M	29.4241219	-98.49362819999999	1409019	Texas
San Diego	BYHfs	M	32.715738	-117.1610838	1355896	California
Dallas	jSYfC	S	32.7766642	-96.79698789999999	1257676	Texas
San Jose	XfDRz	M	37.3382082	-121.8863286	998537	California
Austin	wuyow	S	30.267153	-97.7430608	885400	Texas
Indianapolis	GhoWI	L	39.768403	-86.158068	843393	Indiana

Action Output

```
#  Time      Action                                      Message               Duration / Fetch
1  12:59:03  Apply changes to Assignment2
2  12:59:41  Apply changes to Assignment2                Changes applied
3  13:09:51  select * from exceltable LIMIT 0, 1000      1000 row(s) returned  0.015 sec / 0.000 sec
4  13:10:30  select * from exceltable LIMIT 0, 1000      1000 row(s) returned  0.000 sec / 0.000 sec
5  18:48:27  select * from exceltable LIMIT 0, 1000      1000 row(s) returned  0.000 sec / 0.000 sec
6  18:49:24  select * from JSONTable LIMIT 0, 1000       998 row(s) returned   0.000 sec / 0.015 sec
```

Query Completed

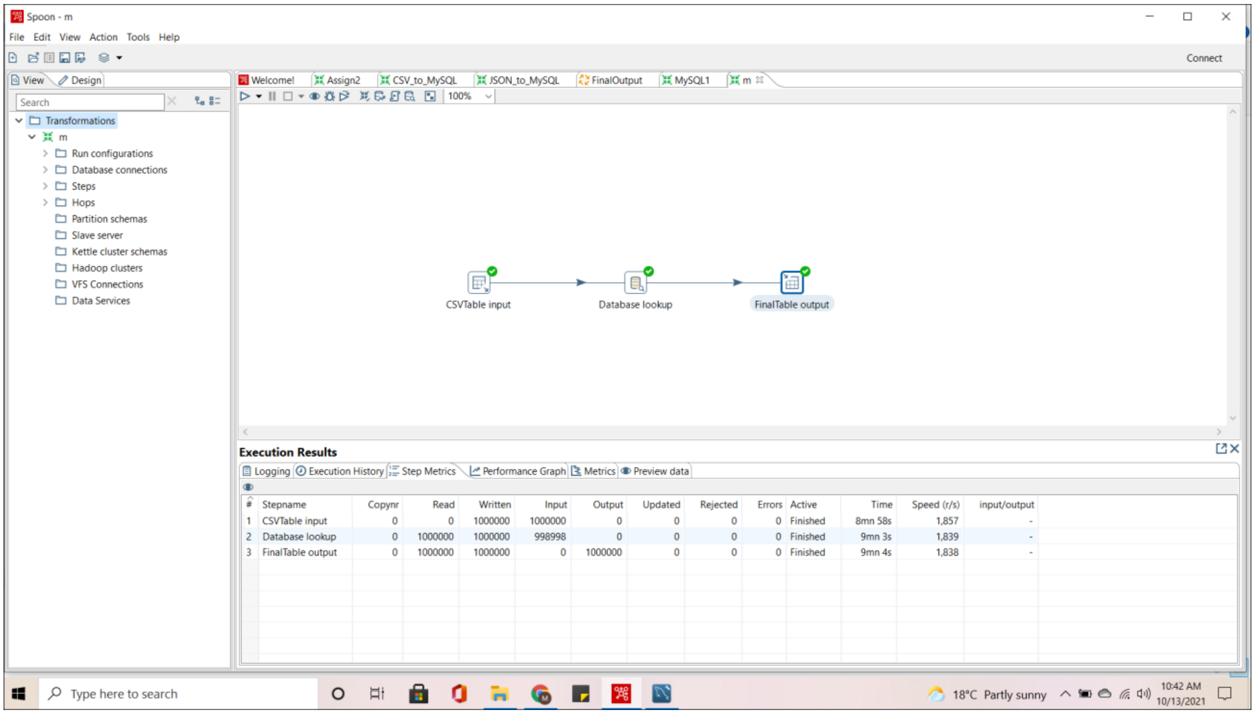
3. Perform Outer Join (or Lookup):



Execution Results

```
2021/10/13 10:40:46 - Database lookup.0 - linenr 950000
2021/10/13 10:40:46 - FinalTable output.0 - linenr 950000
2021/10/13 10:41:08 - CSVTable input.0 - !TableInput.Log.LineNumber!
2021/10/13 10:41:08 - CSVTable input.0 - !TableInput.Log.LineNumber!
2021/10/13 10:41:08 - CSVTable input.0 - !TableInput.Log.FinishedReadingQuery!
2021/10/13 10:41:08 - CSVTable input.0 - Finished processing (I=1000000, O=0, R=0, W=1000000, U=0, E=0)
2021/10/13 10:41:13 - Database lookup.0 - linenr 1000000
2021/10/13 10:41:13 - Database lookup.0 - Finished processing (I=998998, O=0, R=1000000, W=1000000, U=0, E=0)
2021/10/13 10:41:13 - FinalTable output.0 - linenr 1000000
2021/10/13 10:41:13 - FinalTable output.0 - Finished processing (I=0, O=1000000, R=1000000, W=1000000, U=0, E=0)
2021/10/13 10:41:13 - Spoon - The transformation has finished!!
```

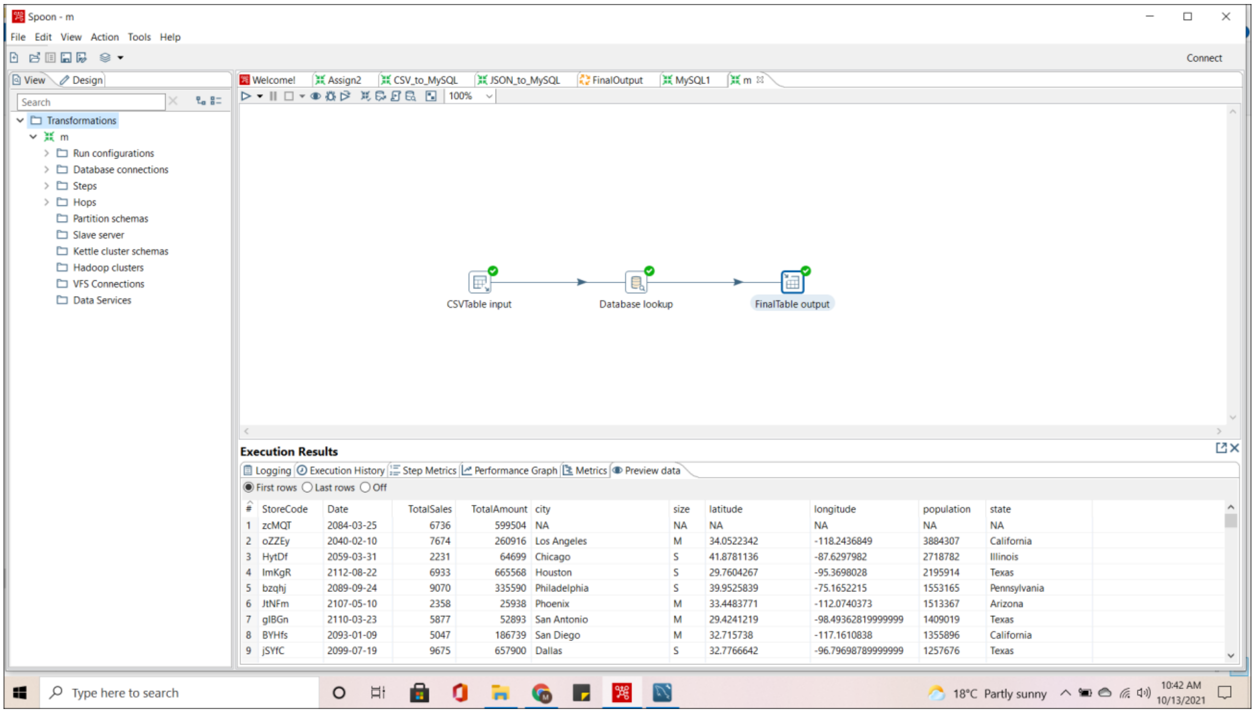
Madhu Jagdale Assignment 2



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	CSVTable input	0	0	1000000	1000000	0	0	0	0	Finished	8mn 58s	1,857	-
2	Database lookup	0	1000000	1000000	998998	0	0	0	0	Finished	9mn 3s	1,839	-
3	FinalTable output	0	1000000	1000000	0	1000000	0	0	0	Finished	9mn 4s	1,838	-

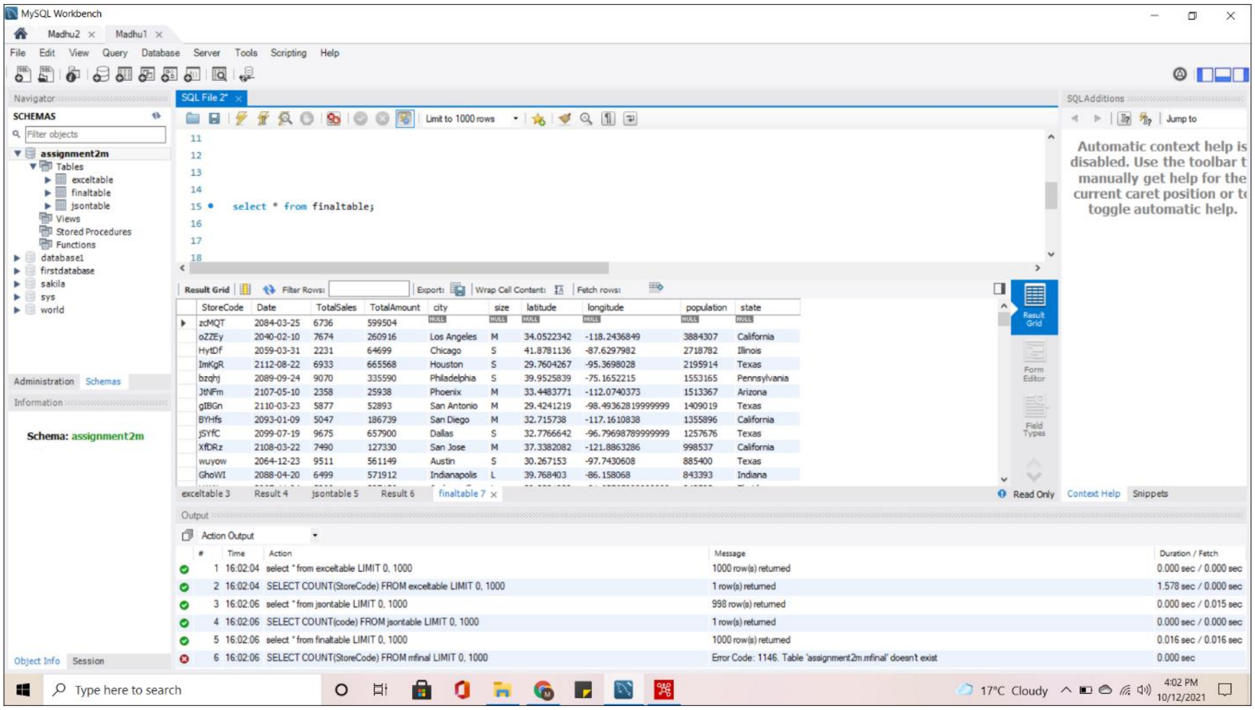


Execution Results

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

● First rows ○ Last rows ○ Off

#	StoreCode	Date	TotalSales	TotalAmount	city	size	latitude	longitude	population	state
1	zcMQT	2084-03-25	6736	599504	NA	NA	NA	NA	NA	NA
2	oZZEy	2040-02-10	7674	260916	Los Angeles	M	34.0522342	-118.2436849	3884307	California
3	HytDf	2059-03-31	2231	64699	Chicago	S	41.8781136	-87.6297982	2718782	Illinois
4	ImKgR	2112-08-22	6933	665568	Houston	S	29.7604267	-95.3698028	2195914	Texas
5	bzqhj	2089-09-24	9070	335590	Philadelphia	S	39.9525839	-75.1652215	1553165	Pennsylvania
6	JtNFm	2107-05-10	2358	25938	Phoenix	M	33.4483771	-112.0740373	1513367	Arizona
7	glBGn	2110-03-23	5877	52893	San Antonio	M	29.4241219	-98.49362819999999	1409019	Texas
8	BYHfs	2093-01-09	5047	186739	San Diego	M	32.715738	-117.1610838	1355896	California
9	jSYfC	2099-07-19	9675	657900	Dallas	S	32.7766642	-96.79698789999999	1257676	Texas

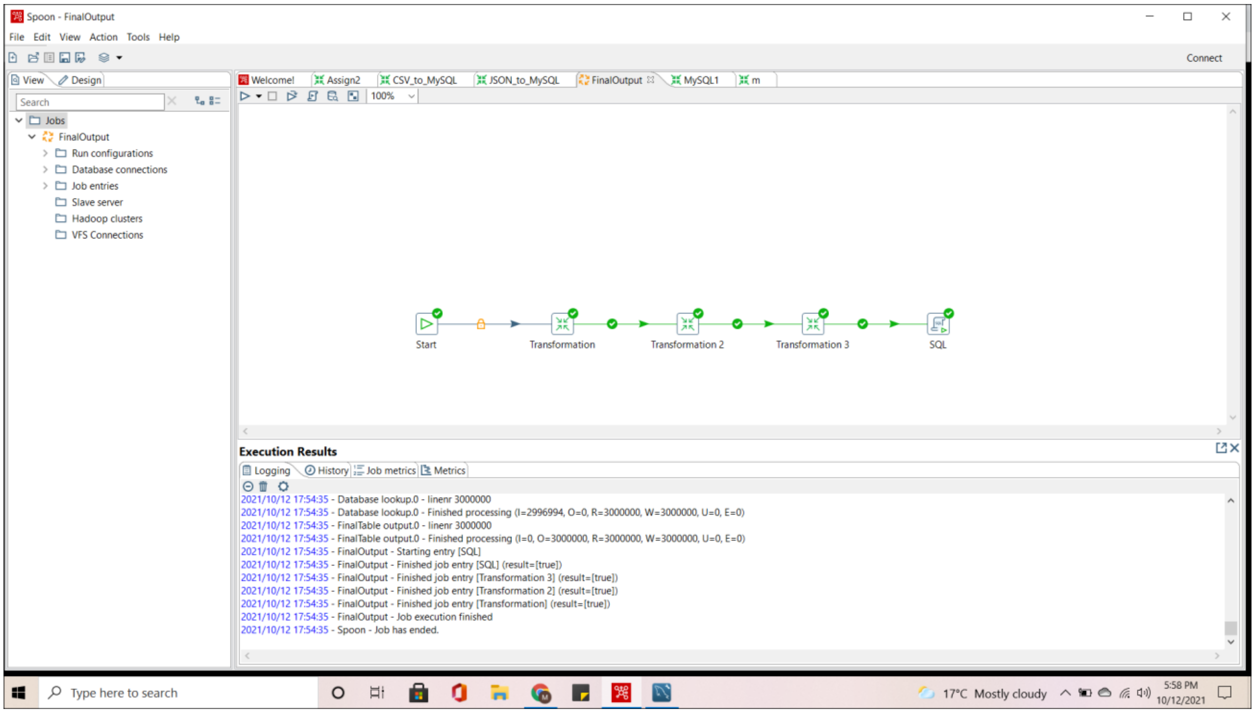


```
select * from finaltable;
```

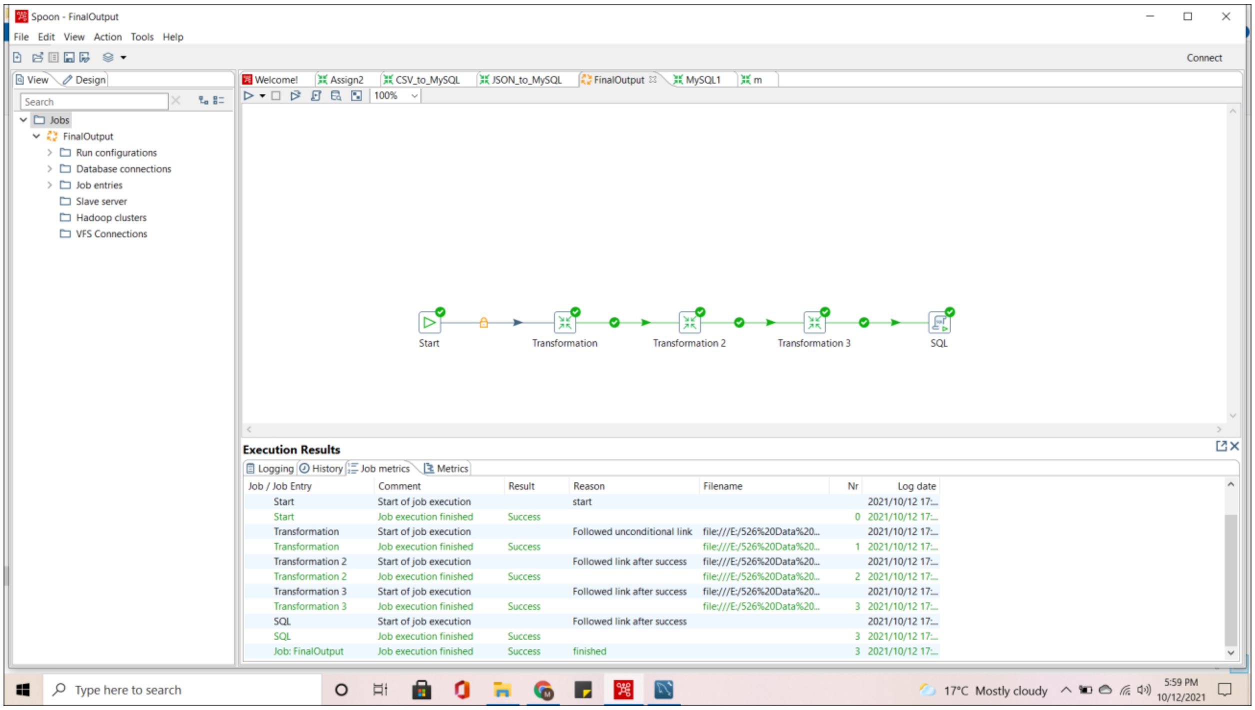
StoreCode	Date	TotalSales	TotalAmount	city	size	latitude	longitude	population	state
zcMQT	2084-03-25	6736	599504	NULL	NULL	NULL	NULL	NULL	NULL
oZZEy	2040-02-10	7674	260916	Los Angeles	M	34.0522342	-118.2436849	3884307	California
HytDf	2059-03-31	2231	64699	Chicago	S	41.8781136	-87.6297982	2718782	Illinois
ImKgR	2112-08-22	6933	665568	Houston	S	29.7604267	-95.3698028	2195914	Texas
bzqhj	2089-09-24	9070	335590	Philadelphia	S	39.9525839	-75.1652215	1553165	Pennsylvania
JtNFm	2107-05-10	2358	25938	Phoenix	M	33.4483771	-112.0740373	1513367	Arizona
gIBGn	2110-03-23	5877	52893	San Antonio	M	29.4241219	-98.49362819999999	1409019	Texas
BYHfs	2093-01-09	5047	186739	San Diego	M	32.715738	-117.1610838	1355896	California
jSYfC	2099-07-19	9675	657900	Dallas	S	32.7766642	-96.79698789999999	1257676	Texas
XfDRz	2108-03-22	7490	127330	San Jose	M	37.3382082	-121.8863286	998537	California
wuyow	2064-12-23	9511	561149	Austin	S	30.267153	-97.7430608	885400	Texas
GhoWI	2088-04-20	6499	571912	Indianapolis	L	39.768403	-86.158068	843393	Indiana

```
#  Time      Action                                                 Message                                                        Duration / Fetch
1  16:02:04  select * from exceltable LIMIT 0, 1000                 1000 row(s) returned                                           0.000 sec / 0.000 sec
2  16:02:04  SELECT COUNT(StoreCode) FROM exceltable LIMIT 0, 1000  1 row(s) returned                                               1.578 sec / 0.000 sec
3  16:02:06  select * from jsontable LIMIT 0, 1000                  998 row(s) returned                                            0.000 sec / 0.015 sec
4  16:02:06  SELECT COUNT(code) FROM jsontable LIMIT 0, 1000        1 row(s) returned                                              0.000 sec / 0.000 sec
5  16:02:06  select * from finaltable LIMIT 0, 1000                 1000 row(s) returned                                           0.016 sec / 0.016 sec
6  16:02:06  SELECT COUNT(StoreCode) FROM mfinal LIMIT 0, 1000      Error Code: 1146. Table 'assignment2m.mfinal' doesn't exist    0.000 sec
```

4. Create a Kettle job to include all the transformations created above.



```
2021/10/12 17:54:35 - Database lookup.0 - linenr 3000000
2021/10/12 17:54:35 - Database lookup.0 - Finished processing (I=2996994, O=0, R=3000000, W=3000000, U=0, E=0)
2021/10/12 17:54:35 - FinalTable output.0 - linenr 3000000
2021/10/12 17:54:35 - FinalTable output.0 - Finished processing (I=0, O=3000000, R=3000000, W=3000000, U=0, E=0)
2021/10/12 17:54:35 - FinalOutput - Starting entry [SQL]
2021/10/12 17:54:35 - FinalOutput - Finished job entry [SQL] (result=[true])
2021/10/12 17:54:35 - FinalOutput - Finished job entry [Transformation 3] (result=[true])
2021/10/12 17:54:35 - FinalOutput - Finished job entry [Transformation 2] (result=[true])
2021/10/12 17:54:35 - FinalOutput - Finished job entry [Transformation] (result=[true])
2021/10/12 17:54:35 - FinalOutput - Job execution finished
2021/10/12 17:54:35 - Spoon - Job has ended.
```

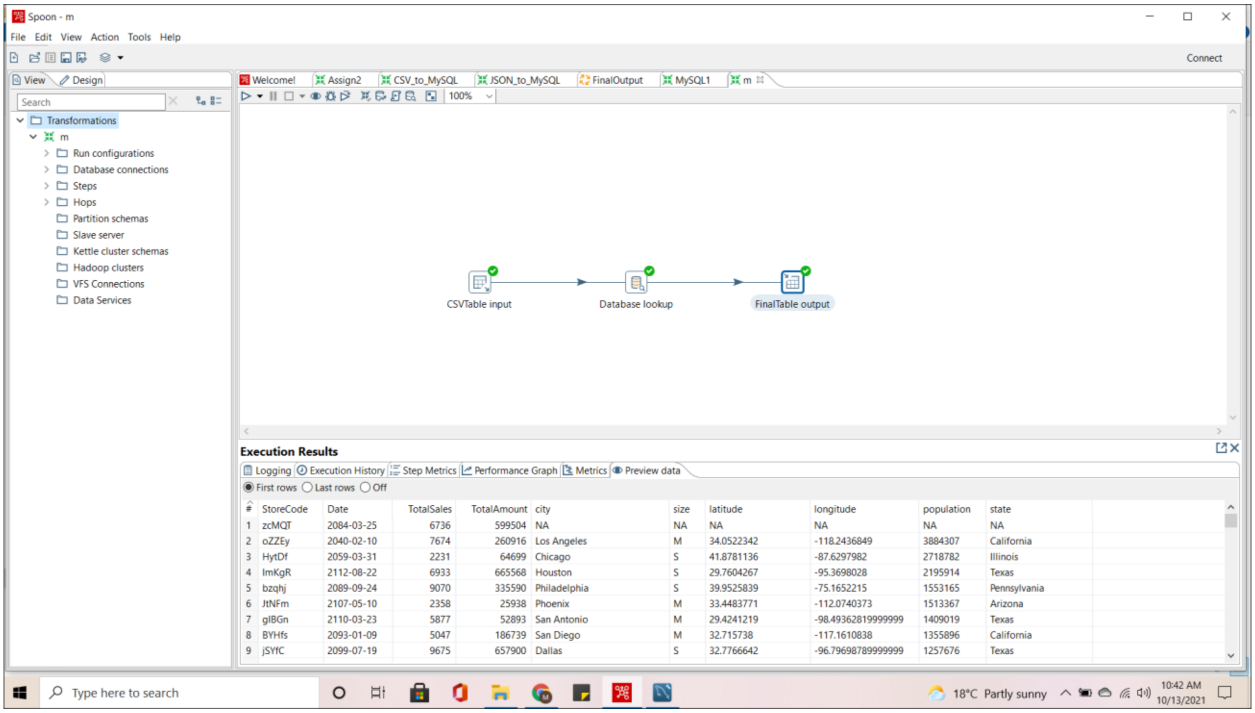


Execution Results

Logging | History | Job metrics | Metrics

Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
Start	Start of job execution		start			2021/10/12 17:...
Start	Job execution finished	Success			0	2021/10/12 17:...
Transformation	Start of job execution		Followed unconditional link	file:///E:/526%20Data%20...		2021/10/12 17:...
Transformation	Job execution finished	Success		file:///E:/526%20Data%20...	1	2021/10/12 17:...
Transformation 2	Start of job execution		Followed link after success	file:///E:/526%20Data%20...		2021/10/12 17:...
Transformation 2	Job execution finished	Success		file:///E:/526%20Data%20...	2	2021/10/12 17:...
Transformation 3	Start of job execution		Followed link after success	file:///E:/526%20Data%20...		2021/10/12 17:...
Transformation 3	Job execution finished	Success		file:///E:/526%20Data%20...	3	2021/10/12 17:...
SQL	Start of job execution		Followed link after success			2021/10/12 17:...
SQL	Job execution finished	Success			3	2021/10/12 17:...
Job: FinalOutput	Job execution finished	Success	finished		3	2021/10/12 17:...

4.a) For those source data rows that do not match a lookup row, set all the column values from the Lookup step to have '**NA**' as their values



Execution Results

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

● First rows ○ Last rows ○ Off

#	StoreCode	Date	TotalSales	TotalAmount	city	size	latitude	longitude	population	state
1	zcMQT	2084-03-25	6736	599504	NA	NA	NA	NA	NA	NA
2	oZZEy	2040-02-10	7674	260916	Los Angeles	M	34.0522342	-118.2436849	3884307	California
3	HytDf	2059-03-31	2231	64699	Chicago	S	41.8781136	-87.6297982	2718782	Illinois
4	ImKgR	2112-08-22	6933	665568	Houston	S	29.7604267	-95.3698028	2195914	Texas
5	bzqhj	2089-09-24	9070	335590	Philadelphia	S	39.9525839	-75.1652215	1553165	Pennsylvania
6	JtNFm	2107-05-10	2358	25938	Phoenix	M	33.4483771	-112.0740373	1513367	Arizona
7	gIBGn	2110-03-23	5877	52893	San Antonio	M	29.4241219	-98.49362819999999	1409019	Texas
8	BYHfs	2093-01-09	5047	186739	San Diego	M	32.715738	-117.1610838	1355896	California
9	jSYfC	2099-07-19	9675	657900	Dallas	S	32.7766642	-96.79698789999999	1257676	Texas