

Jignasuben R Vekariya- 002981797

Homework 6

As an individual dataset contains various features and millions of records. After reading the commitment to protecting nursing home and residents, it's crucial thing to have track of vaccination background of residents and staff.

Therefore, I trimmed dataset and plan to load data only about vaccination details like either they're vaccinated or not, if vaccinated which vaccine they received and how many dose.

QU1: How loading data? Using SSIS or BCP

From all provided datasets, I selected latest dataset which contain details up to March 2022 because I want to analyze the vaccinated staff and residents and as we know vaccination is mostly increase after 2021 and until 2022 booster requirement is also become necessary.

Step1: Extract data

Extract vaccination data columns from the dataset and create whole new dataset.

Step2: Filter data

Replace column values which contains delimiter comma(,) to space in the string. As it'll help further in SSIS process while mapping the columns.

Step3: Loading data

After extracting and filtering data, load the data into SQL Server using SSIS and BCP process.

QU2: Load data into SQL using SSIS (10 Points)

The screenshot shows the SQL Server Data Tools (SSDT) interface. On the left, the Data Flow Task is visible in the Design view, showing a 'Flat File Source' connected to an 'OLE DB Destination'. The right pane displays the 'Flat File Source Output Data Viewer at Data Flow Task'. The data is presented in a table with the following columns: Week Ending, Federal Provider Number, Provider Name, Provider City, Provider State, and Provider ID. The data is sorted by Week Ending in descending order.

Week Ending	Federal Provider Number	Provider Name	Provider City	Provider State	Provider ID
12/12/2021	15019	MERRY WOOD LODGE	ELMORE	AL	33456
12/19/2021	15019	MERRY WOOD LODGE	ELMORE	AL	33456
12/26/2021	15019	MERRY WOOD LODGE	ELMORE	AL	33456
1/2/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
1/9/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
1/16/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
1/23/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
1/30/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
2/6/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
2/13/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
2/20/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
2/27/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
3/6/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
3/13/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
3/20/2022	15019	MERRY WOOD LODGE	ELMORE	AL	33456
5/2/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
5/9/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
6/7/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
6/14/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
6/21/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
6/28/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
7/5/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
7/12/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
7/19/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
7/26/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575
8/2/2022	15023	HATLEY HEALTH CARE INC	CLANTON	AL	20575

QU3: Load data using BCP (5 points)

The screenshot shows a Windows Command Prompt window with the following commands and output:

```
error; see previous errors for some of these.

C:\tmp>bcp c19 in c:\tmp\c19_M2022.csv -d hw6 -T -S LAPTOP-IV8MQ3U3\SQL2019 -f hw6bcp.fmt
SQLState = S0002, NativeError = 208
Error = [Microsoft][ODBC Driver 17 for SQL Server][SQL Server]Invalid object name 'c19'.
SQLState = 37000, NativeError = 11529
Error = [Microsoft][ODBC Driver 17 for SQL Server][SQL Server]The metadata could not be determined because every code path results in an
error; see previous errors for some of these.

C:\tmp>bcp c19 in c:\tmp\c19_M2022.csv -T -S LAPTOP-IV8MQ3U3\SQL2019 -d hw6 -f hw6bcp.fmt
SQLState = S0002, NativeError = 208
Error = [Microsoft][ODBC Driver 17 for SQL Server][SQL Server]Invalid object name 'c19'.
SQLState = 37000, NativeError = 11529
Error = [Microsoft][ODBC Driver 17 for SQL Server][SQL Server]The metadata could not be determined because every code path results in an
error; see previous errors for some of these.

C:\tmp>bcp c19 in c:\tmp\c19_M2022.csv -T -S LAPTOP-IV8MQ3U3\SQL2019 -d hw6 -f hw6bcp.fmt
SQLState = HY000, NativeError = 0
Error = [Microsoft][ODBC Driver 17 for SQL Server][SQL Server]Syntax error at line 1 column 0 in xml format file.

C:\tmp>bcp c19 in c:\tmp\c19_M2022.csv -T -S LAPTOP-IV8MQ3U3\SQL2019 -c -d hw6 -f hw6_bcp.fmt
Warning: -f overrides -c.
SQLState = HY000, NativeError = 0
Error = [Microsoft][ODBC Driver 17 for SQL Server][SQL Server]Syntax error at line 1 column 0 in xml format file.

C:\tmp>bcp c19 in c:\tmp\c19_M2022.csv -T -S LAPTOP-IV8MQ3U3\SQL2019 -c -d hw6 -F 1
Starting copy...
0 rows copied.
Network packet size (bytes): 4096
Clock Time (ms.) Total : 281

C:\tmp>
```

QU4: Analyze data (30 Points). Provide set of analysis (up to 1 page/2 table or graph)

Analysis:1

The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains the following SQL code:

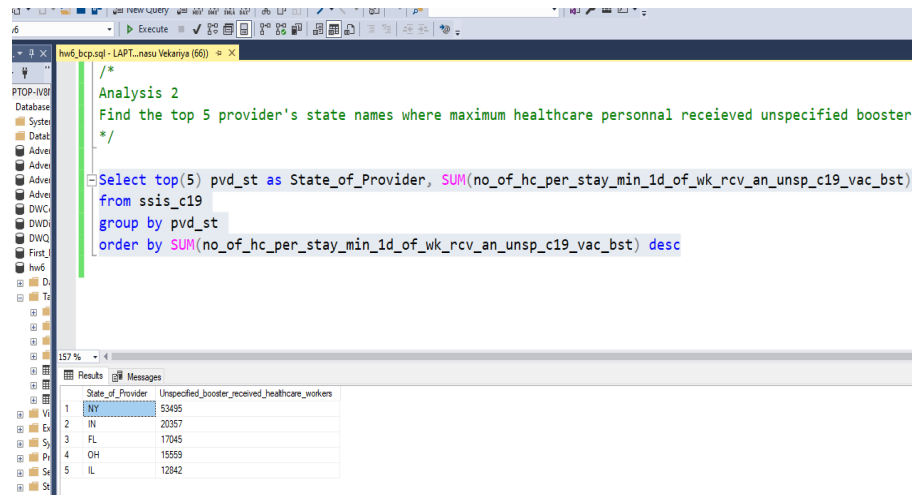
```
/*
Analysis 1
Find the County name where percentage of fully vaccinated resident is less than 70%
*/
Select county, SUM(no_of_res_stay_min_id_of_wk) as total_stayed_resident_for_this_week, SUM(no_of_res_stay_min_id_of_wk_rcv_comp_c19_vac) as vaccinated_resident_
from ssis_c19
group by county

select county, (vaccinated_resident_for_this_week * 100)/total_stayed_resident_for_this_week as percentage_of_vaccinated_resident
from analysis1
where vaccinated_resident_for_this_week != 0 AND ((vaccinated_resident_for_this_week * 100)/total_stayed_resident_for_this_week)<70
/* for above query, we have to put where condition because in caribou country 0 residents stayed over the time. */
```

The Results pane displays the following data:

county	percentage_of_vaccinated_resident
1 Yavapai	67
2 Pasquotank	69
3 St. Lucie	67
4 Ben Hill	53
5 Santa Rosa	69
6 Fannin	62
7 Highlands	65
8 Hooking	55
9 Doniphan	69
10 Wilbaur	66
11 McCreary	65
12 Habersham	68
13 McCurtain	65
14 Gates	55
15 Mohave	68
16 Valencia	66
17 Sevier	60
18 Hardee	60
19 Beaver	62
20 Tehama	69
21 Bacon	67

Analysis :2

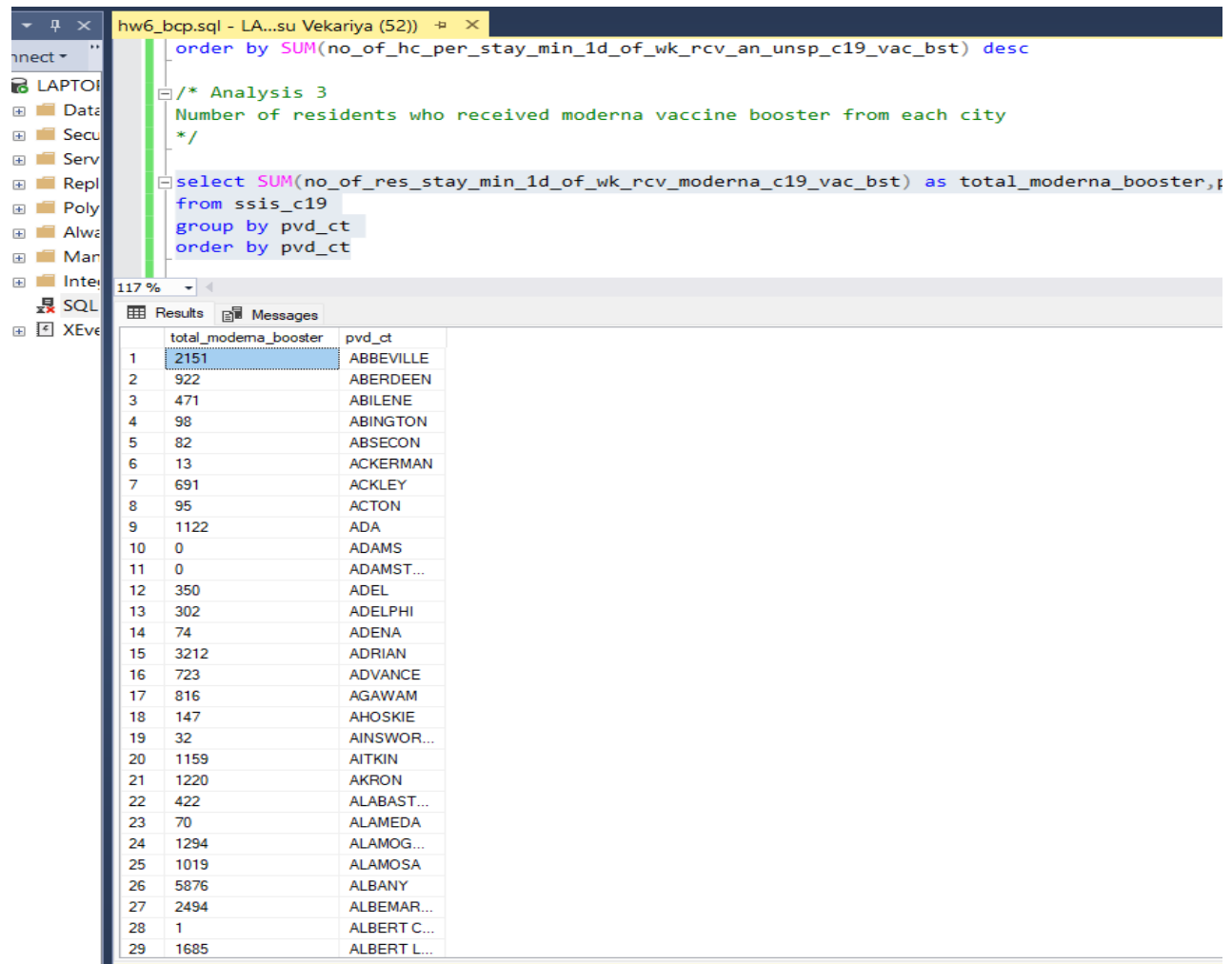


```
/*
Analysis 2
Find the top 5 provider's state names where maximum healthcare personnel received unspecified booster
*/

Select top(5) pvd_st as State_of_Provider, SUM(no_of_hc_per_stay_min_1d_of_wk_rcv_an_unsp_c19_vac_bst)
from ssis_c19
group by pvd_st
order by SUM(no_of_hc_per_stay_min_1d_of_wk_rcv_an_unsp_c19_vac_bst) desc
```

State_of_Provider	Unspecified_booster_received_healthcare_workers
NY	53495
WI	20357
FL	17045
OH	15569
IL	12842

Analysis :3



```
order by SUM(no_of_hc_per_stay_min_1d_of_wk_rcv_an_unsp_c19_vac_bst) desc

/* Analysis 3
Number of residents who received moderna vaccine booster from each city
*/

select SUM(no_of_res_stay_min_1d_of_wk_rcv_moderna_c19_vac_bst) as total_moderna_booster,
from ssis_c19
group by pvd_ct
order by pvd_ct
```

total_moderna_booster	pvd_ct
2151	ABBEVILLE
922	ABERDEEN
471	ABILENE
98	ABINGTON
82	ABSECON
13	ACKERMAN
691	ACKLEY
95	ACTON
1122	ADA
0	ADAMS
0	ADAMST...
350	ADEL
302	ADELPHI
74	ADENA
3212	ADRIAN
723	ADVANCE
816	AGAWAM
147	AHOSKIE
32	AINSWOR...
1159	AITKIN
1220	AKRON
422	ALABAST...
70	ALAMEDA
1294	ALAMOG...
1019	ALAMOSA
5876	ALBANY
2494	ALBEMAR...
1	ALBERT C...
1685	ALBERT L...

Analysis:4

The screenshot shows a SQL Server Enterprise Manager interface. The query editor displays a query for Analysis 4. The query is as follows:

```
/* Analysis 4
Retrive the total number of vaccinated residents and with unknown vaccination status
*/
select SUM(no_of_res_stay_min_1d_of_wk) as res_stay_minimum_1day,
(SUM(no_of_res_stay_min_1d_of_wk)-SUM(no_of_res_stay_min_1d_of_wk_with_an_ukwn_c19_vac_Status)) as vaccinated_resident,
SUM(no_of_res_stay_min_1d_of_wk_with_an_ukwn_c19_vac_Status) as Unknown_vaccination_status
from ssis_c19
```

The Results pane shows the following data:

res_stay_minimum_1day	vaccinated_resident	Unknown_vaccination_status
35854766	35311866	542900

Analysis:5

The screenshot shows a SQL Server Enterprise Manager interface. The query editor displays a query for Analysis 5. The query is as follows:

```
/* Analysis 5
retrive the number of health workers who are fully or partial vaccinated and/or having medicale contradiction or decline to take vaccine
*/
select SUM(no_of_hc_per_Eligible_to_Work_min_1d_of_wk_rcv_comp_c19_vac) as fully_vaccinated_healthcare_workers,
SUM(no_of_hc_per_Eligible_to_Work_min_1d_of_wk_rcv_Partial_c19_vac) as partial_vaccinated_healthcare_workers,
SUM(no_of_hc_per_Eligible_to_Work_min_1d_of_wk_with_med_contraindication_to_c19_vac) as healthcare_workers_have_medical_contradiction,
SUM(no_of_hc_per_Eligible_to_Work_min_1d_of_wk_decline_c19_vac) as healthcare_workers_who_decline_vaccination,
pvd_st as States
from ssis_c19
group by pvd_st
order by pvd_st
```

The Results pane shows the following data:

fully_vaccinated_healthcare_workers	partial_vaccinated_healthcare_workers	healthcare_workers_have_medical_contradiction	healthcare_workers_who_decline_vaccination	States
72064	1405	511	10913	AK
786153	48760	7849	236907	AL
665458	39229	9692	126521	AR
528421	19621	9019	112647	AZ
2903540	39651	27272	182747	CA
820619	16948	11153	82210	CO
1394883	25862	14320	70452	CT
131275	4714	324	4318	DC
224909	7250	1234	35633	DE
2683447	133703	28819	879042	FL
1029355	58539	8923	271871	GA
267645	2494	763	11332	HI
1116538	30986	13560	281678	IA
190666	7174	4773	64249	ID
2499541	78241	27806	475895	IL
1403702	66812	34549	511863	IN
796117	26938	10897	231553	KS
826896	48039	9918	290692	KY
729155	59715	6616	268173	LA
1916792	33766	11659	103588	MA
1167361	34570	8147	121411	MD
386777	5902	2349	32594	ME
1463438	72564	20741	497772	MI
1469382	34144	11929	304566	MN
1083285	61521	16431	445273	MO
556339	36629	6051	194501	MS
142924	3867	2091	50594	MT

Analysis:6

The screenshot shows a SQL Server Enterprise Manager window with a query titled "hw6_bcp.sql - LA...su Vekariya (52)". The query is as follows:

```
/* Analysis 6
Retrieve the provider name where more than 50% residents are decline to take vaccines
*/
select pvd_Name as provider_name
from ssis_c19
where no_of_res_stay_min_1d_of_wk_decline_c19_vac != 0 AND ((no_of_res_stay_min_1d_of_wk_decline_c19_vac * 100)/no_of_r
group by pvd_Name
```

The query results are displayed in a table with the following data:

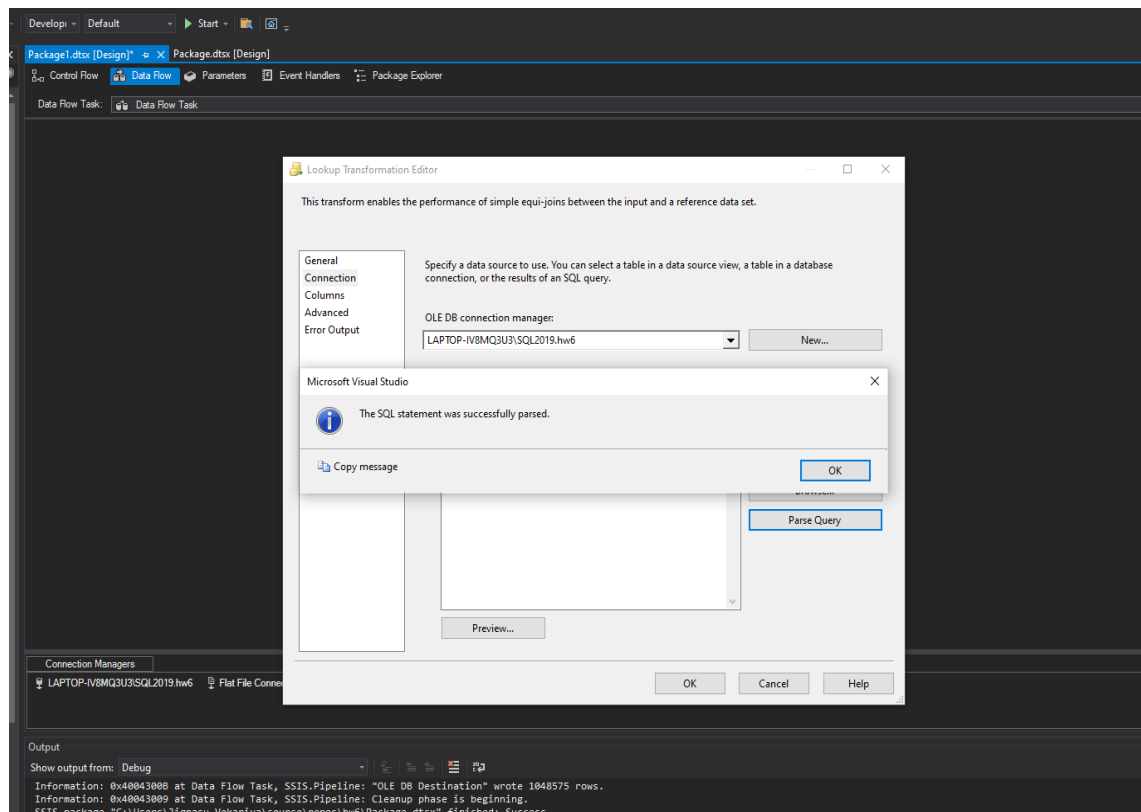
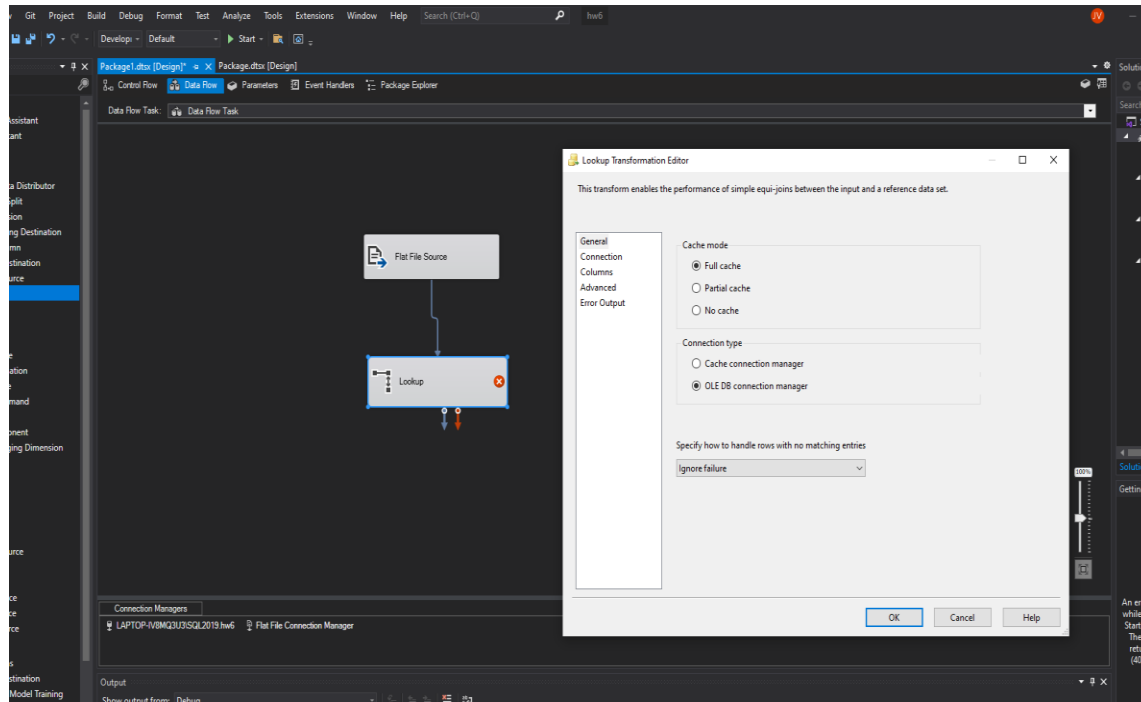
provider_name
COUNTRY VILLA EAST NURSING CENTER
LAURELWOOD COMMUNITY LIVING CENTER
WELLBRIDGE OF NOVI LLC
AERIUS HEALTH CENTER
GOVERNORS CREEK HEALTH AND REHABILITATION
FOURAKER HILLS REHAB AND NURSING CENTER
LEE COUNTY CARE AND REHABILITATION CENTER
DIVINE REHABILITATION AND NURSING AT CANAL POINTE
DIVERSICARE OF BRADFORD PLACE
LIFE CARE CENTER OF SOUTH LAS VEGAS
NORWOOD TOWERS POST-ACUTE
NSPIRE HEALTHCARE TAMARAC
BUCKEYE TERRACE REHABILITATION AND NURSING CENTER
LAKE CROSSING HEALTH CENTER
AVANTE AT LAKE WORTH INC
BIG CREEK NURSING AND REHABILITATION COMMUNITY
LAKESIDE MANOR NURSING AND REHABILITATION CENTER
HEALTH AND REHABILITATION CENTRE AT DOLPHINS VIEW
HOMESTEAD HEALTHCARE CENTER
ADVANCED HEALTH CARE OF LAS VEGAS
VERRAZANO NURSING AND POST-ACUTE CENTER
ACCORDIUS HEALTH AT WILMINGTON
SUMMIT'S TRACE HEALTHCARE CENTER
SPRENGER HEALTH CARE OF MISHAWAKA
NORTH DADE NURSING AND REHABILITATION CENTER
VALPARAISO CARE & REHABILITATION
NORTHSHORE EXTENDED CARE HOSPITAL

QU5: What's step to take if file provided again? What step for loading the data. (10 Points)

First, I try to load data using BCP which I was unable to perform because of some error.

Further for the new data, I'll update using incremental load in SSIS which compares source table and target table based on date or time.

QU6: Modify SSIS package to handle the next month data



As I loaded recent data in first SSIS, I just uploaded process of incremental load.

QU7: The file will also be in module 5

Create new file of test data show.. how created the input file (5 Points)