Final Project Part 2

Sangzun Park

1. Create Conceptual Diagram/Schema for database

The original CSV file was a table with 52 columns and 4498 rows. I analyzed 52 columns, and through this analysis I categorized each column. Categorized columns were largely classified into three categories.

1. Company Name.
2. Medicare drug Name
3. Year & Data(Claim, Spending, beneficiary, etc..)

2011	2012	2013	2014	2015
Data(Numeric)				

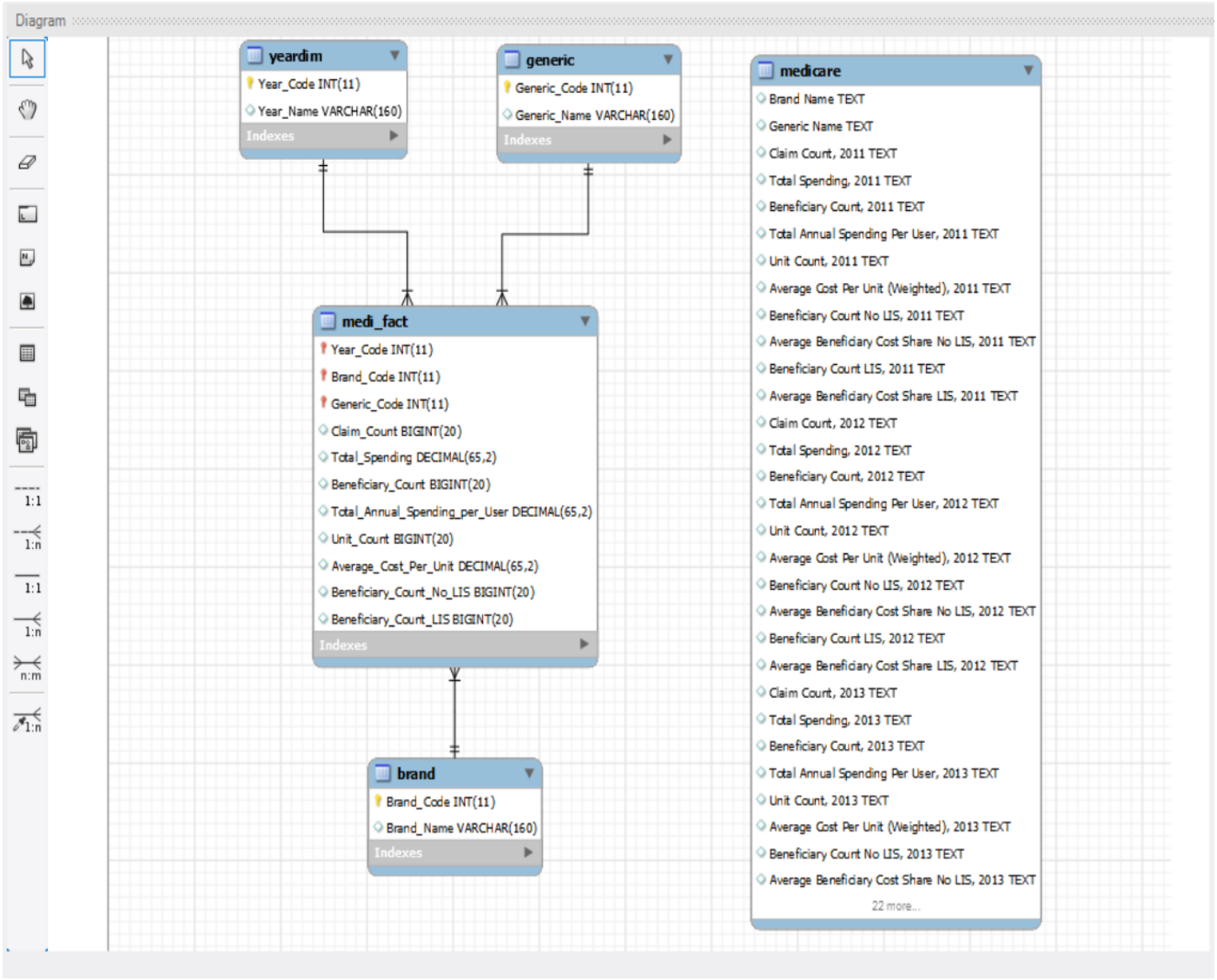
At the same time, I classified each type of data as follows.

1. Company Name : Text(Char)
2. Medicare drug Name : Text(Char)
3. Year & Data(Claim, Spending, beneficiary, etc..) : Integer and Decimal

Based on this analysis, I designed the entire table to resemble a Star Schema structure.

Fact Table	Dimension_1	Dimension_2	Dimension_3
Medifact(Numeric)	Brand	Generic	year

Finally, I have completed the structure, diagram of the whole database as below.



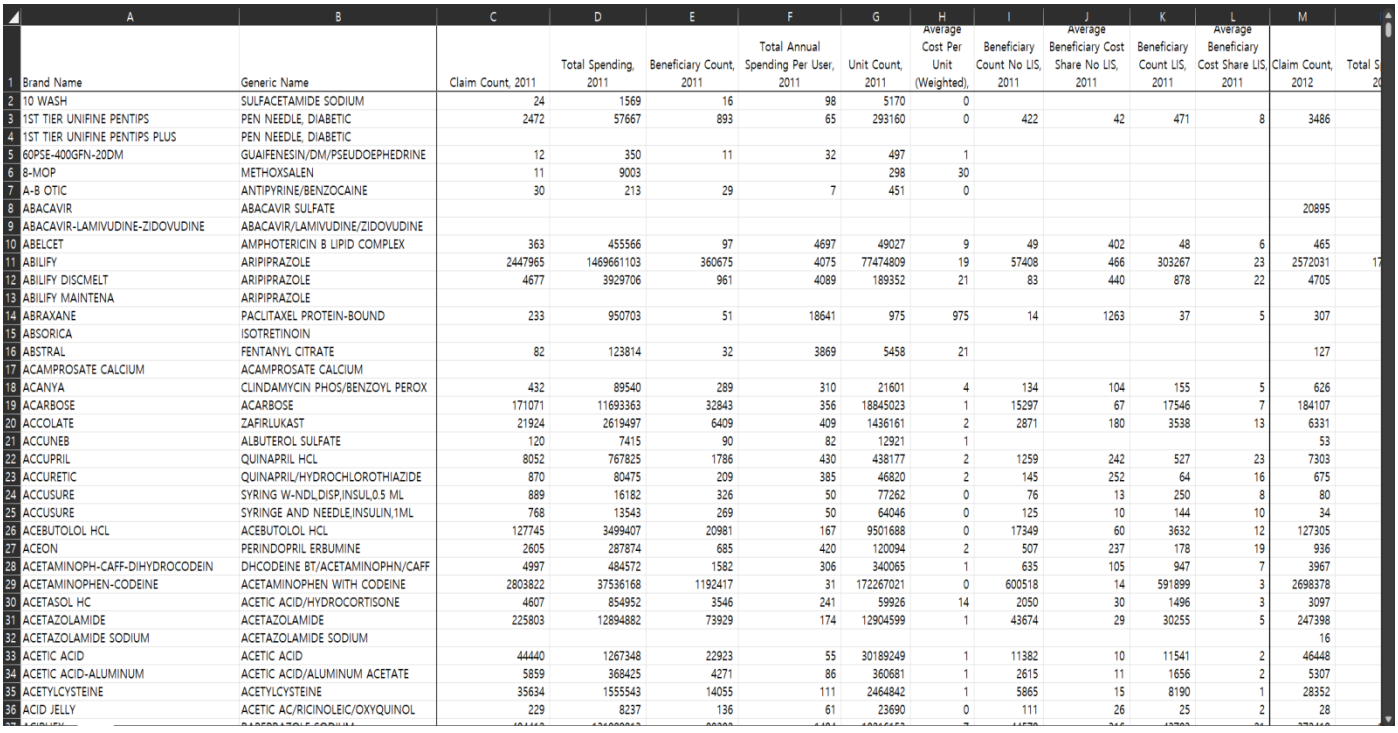
(Left : After Design / Right : Before Design)

2. Database

Preprocessing on Exel&CSV data

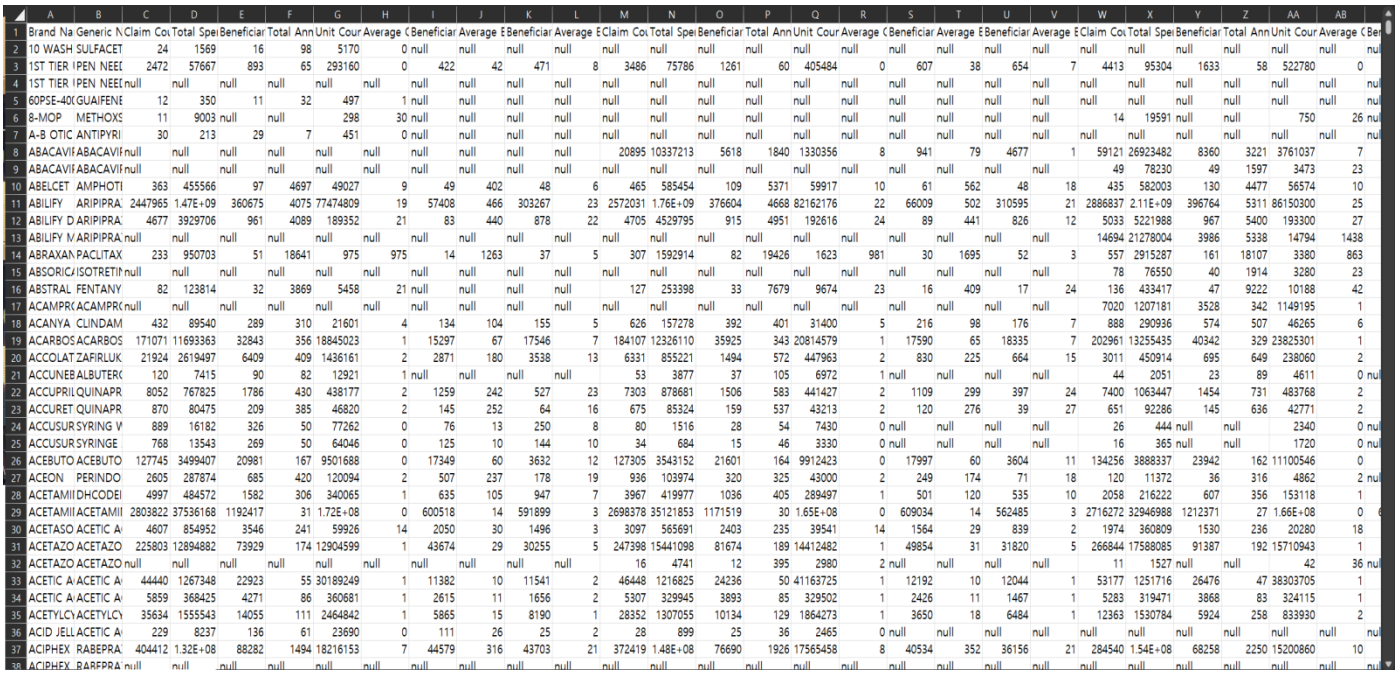
Regarding blank items in the original CSV, Mysql deleted all corresponding rows during the import process. Therefore, I changed all blank values to null values in the CSV file in advance and imported them into Mysql. In addition, the data type of each column in the CSV file has been modified so that Mysql can recognize it.

[Before]



	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	Brand Name	Generic Name	Claim Count, 2011	Total Spending, 2011	Beneficiary Count, 2011	Total Annual Spending Per User, 2011	Unit Count, 2011	Average Cost Per Unit (Weighted),	Beneficiary Count No LIS, 2011	Average Beneficiary Cost Share No LIS, 2011	Beneficiary Count LIS, 2011	Average Beneficiary Cost Share LIS, 2011	Claim Count, 2012	Total S
2	10 WASH	SULFACETAMIDE SODIUM	24	1569	16	98	5170	0						
3	1ST TIER UNIFINE PENTIPS	PEN NEEDLE, DIABETIC	2472	57667	893	65	293160	0	422	42	471	8	3486	
4	1ST TIER UNIFINE PENTIPS PLUS	PEN NEEDLE, DIABETIC												
5	60PSE-400GFN-20DM	GUAIFENESIN/DM/PSEUDOEPHEDRINE	12	350	11	32	497	1						
6	8-MOP	METHOXSALEN	11	9003			298	30						
7	A-B OTIC	ANTIPYRINE/BENZOCAINE	30	213	29	7	451	0						
8	ABACAVIR	ABACAVIR SULFATE											20895	
9	ABACAVIR-LAMIVUDINE-ZIDOVUDINE	ABACAVIR/LAMIVUDINE/ZIDOVUDINE												
10	ABELCET	AMPHOTERICIN B LIPID COMPLEX	363	455566	97	4697	49027	9	49	402	48	6	465	
11	ABILIFY	ARIPIPRAZOLE	2447965	1469661103	360675	4075	77474809	19	57408	466	303267	23	2572031	17
12	ABILIFY DISCMELT	ARIPIPRAZOLE	4677	3929706	961	4089	189352	21	83	440	878	22	4705	
13	ABILIFY MAINTENA	ARIPIPRAZOLE												
14	ABRAXANE	PACLITAXEL PROTEIN-BOUND	233	950703	51	18641	975	975	14	1263	37	5	307	
15	ABSORICA	ISOTRETINOIN												
16	ABSTRAL	FENTANYL CITRATE	82	123814	32	3869	5458	21					127	
17	ACAMPROSATE CALCIUM	ACAMPROSATE CALCIUM												
18	ACANYA	CLINDAMYCIN PHOS/BENZOYL PEROX	432	89540	289	310	21601	4	134	104	155	5	626	
19	ACARBOSE	ACARBOSE	171071	11693363	32843	356	18845023	1	15297	67	17546	7	184107	
20	ACCOLATE	ZAFIRLUKAST	21924	2619497	6409	409	1436161	2	2871	180	3538	13	6331	
21	ACCUNEB	ALBUTEROL SULFATE	120	7415	90	82	12921	1					53	
22	ACCUPRIL	QUINAPRIL HCL	8052	767825	1786	430	438177	2	1259	242	527	23	7303	
23	ACCURETIC	QUINAPRIL/HYDROCHLOROTHIAZIDE	870	80475	209	385	46820	2	145	252	64	16	675	
24	ACCUSURE	SYRING W-NDL,DISP,INSUL,0.5 ML	889	16182	326	50	77262	0	76	13	250	8	80	
25	ACCUSURE	SYRINGE AND NEEDLE,INSULIN,1ML	768	13543	269	50	64046	0	125	10	144	10	34	
26	ACEBUTOLOL HCL	ACEBUTOLOL HCL	127745	3499407	20981	167	9501688	0	17349	60	3632	12	127305	
27	ACEON	PERINDOPRIL ERBUMINE	2605	287874	685	420	120094	2	507	237	178	19	936	
28	ACETAMINOPH-CAFF-DIHYDROCODEIN	DHCODEINE BT/ACETAMINOPHN/CAFF	4997	484572	1582	306	340065	1	635	105	947	7	3967	
29	ACETAMINOPHEN-CODEINE	ACETAMINOPHEN WITH CODEINE	2803822	37536168	1192417	31	172267021	0	600518	14	591899	3	2698378	
30	ACETASOL HC	ACETIC ACID/HYDROCORTISONE	4607	854952	3546	241	59926	14	2050	30	1496	3	3097	
31	ACETAZOLAMIDE	ACETAZOLAMIDE	225803	12894882	73929	174	12904599	1	43674	29	30255	5	247398	
32	ACETAZOLAMIDE SODIUM	ACETAZOLAMIDE SODIUM											16	
33	ACETIC ACID	ACETIC ACID	44440	1267348	22923	55	30189249	1	11382	10	11541	2	46448	
34	ACETIC ACID-ALUMINUM	ACETIC ACID/ALUMINUM ACETATE	5859	368425	4271	86	360681	1	2615	11	1656	2	5307	
35	ACETYLCYSTEINE	ACETYLCYSTEINE	35634	1555543	14055	111	2464842	1	5865	15	8190	1	28352	
36	ACID JELLY	ACETIC AC/RICINOLEIC/OXYQUINOL	229	8237	136	61	23690	0	111	26	25	2	28	

[After]



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	
1	Brand Na	Generic N	Claim Cou	Total Spei	Beneficiar	Total Ann	Unit Cour	Average C	Beneficiar	Average E	Beneficiar	Average E	Claim Cou	Total Spei	Beneficiar	Total Ann	Unit Cour	Average C	Beneficiar	Average E	Beneficiar	Average E	Claim Cou	Total Spei	Beneficiar	Total Ann	Unit Cour	Average C	Ber
2	10 WASH	SULFACET	24	1569	16	98	5170	0	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	nul
3	1ST TIER	PEN NEEI	2472	57667	893	65	293160	0	422	42	471	8	3486	75786	1261	60	405484	0	607	38	654	7	4413	95304	1633	58	522780	0	
4	1ST TIER	PEN NEEI	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	nul
5	60PSE-40(	GUAIFENE	12	350	11	32	497	1	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	nul
6	8-MOP	METHOXS	11	9003	null	null	298	30	null	null	null	null	null	null	null	null	null	null	null	null	null	null	14	19591	null	null	750	26	nul
7	A-B OTIC	ANTIPYRI	30	213	29	7	451	0	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	nul
8	ABACAVIF	ABACAVIF	null	null	null	null	null	null	null	null	null	null	20895	10337213	5618	1840	1330356	8	941	79	4677	1	59121	26923482	8360	3221	3761037	7	
9	ABACAVIF	ABACAVIF	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	49	78230	49	1597	3473	23	
10	ABELCET	AMPHOTI	363	455566	97	4697	49027	9	49	402	48	6	465	585454	109	5371	59917	10	61	562	48	18	435	582003	130	4477	56574	10	
11	ABILIFY	ARIPIPRA	2447965	1.47E+09	360675	4075	77474809	19	57408	466	303267	23	2572031	1.76E+09	376604	4668	82162176	22	66009	502	310595	21	2886837	2.11E+09	396764	5311	86150300	25	
12	ABILIFY D	ARIPIPRA	4677	3929706	961	4089	189352	21	83	440	878	22	4705	4529795	915	4951	192616	24	89	441	826	12	5033	5221988	967	5400	193300	27	
13	ABILIFY M	ARIPIPRA	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	14694	21278004	3986	5338	14794	1438	
14	ABRAXAN	PACLITAX	233	950703	51	18641	975	975	14	1263	37	5	307	1592914	82	19426	1623	981	30	1695	52	3	557	2915287	161	18107	3380	863	
15	ABSORICA	ISOTRETI	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	78	76550	40	1914	3280	23	
16	ABSTRAL	FENTANY	82	123814	32	3869	5458	21	null	null	null	null	127	253398	33	7679	9674	23	16	409	17	24	136	433417	47	9222	10188	42	
17	ACAMPRO	ACAMPRO	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	7020	1207181	3528	342	1149195	1	
18	ACANYA	CLINDAM	432	89540	289	310	21601	4	134	104	155	5	626	157278	392	401	31400	5	216	98	176	7	888	290936	574	507	46265	6	
19	ACARBOS	ACARBOS	171071	11693363	32843	356	18845023	1	15297	67	17546	7	184107	12326110	35925	343	20814579	1	17590	65	18335	7	202961	13255435	40342	329	23825301	1	
20	ACCOLAT	ZAFIRLUK	21924	2619497	6409	409	1436161	2	2871	180	3538	13	6331	855221	1494	572	447963	2	830	225	664	15	3011	450914	695	649	238060	2	
21	ACCUNEB	ALBUTER(	120	7415	90	82	12921	1	null	null	null	null	53	3877	37	105	6972	1	null	null	null	null	44	2051	23	89	4611	0	nul
22	ACCUPRIL	QUINAPR	8052	767825	1786	430	438177	2	1259	242	527	23	7303	878681	1506	583	441427	2	1109	299	397	24	7400	1063447	1454	731	483768	2	
23	ACCURET	QUINAPR	870	80475	209	385	46820	2	145	252	64	16	675	85324	159	537	43213	2	120	276	39	27	651	92286	145	636	42771	2	
24	ACCUSUR	SYRING W	889	16182	326	50	77262	0	76	13	250	8	80	1516	28	54	7430	0	null	null	null	null	26	444	null	null	2340	0	nul
25	ACCUSUR	SYRINGE	768	13543	269	50	64046	0	125	10	144	10	34	684	15	46	3330	0	null	null	null	null	16	365	null	null	1720	0	nul
26	ACEBUTO	ACEBUTO	127745	3499407	20981	167	9501688	0	17349	60	3632	12	127305	3543152	21601	164	9912423	0	17997	60	3604	11	134256	3888337	23942	162	11100546	0	
27	ACEON	PERINDO	2605	287874	685	420	120094	2	507	237	178	19	936	103974	320	325	43000	2	249	174	71	18	120	11372	36	316	4862	2	nul
28	ACETAMIN	DHCODEI	4997	484572	1582	306	340065	1	635	105	947	7	3967	419977	1036	405	289497	1	501	120	535	10	2058	216222	607	356	153118	1	
29	ACETAMIN	ACETAMIN	2803822	37536168	1192417	31	1.72E+08	0	600518	14	591899	3	2698378	35121853	1171519	30	1.65E+08	0	609034	14	562485	3	2716272	32946988	1212371	27	1.66E+08	0	
30	ACETASO	ACETIC A	4607	854952	3546	241	59926	14	2050	30	1496	3	3097	565691	2403	235	39541	14	1564	29	839	2	1974	360809	1530	236	20280	18	
31	ACETAZO	ACETAZO	225803	12894882	73929	174	12904599	1	43674	29	30255	5	247398	15441098	81674	189	14412482	1	49854	31	31820	5	266844	17588085	91387	192	15710943	1	
32	ACETAZO	ACETAZO	null	null	null	null	null	null	null	null	null	null	16	4741	12	395	2980	2	null	null	null	null	11	1527	null	null	42	36	nul
33	ACETIC A	ACETIC A	44440	1267348	22923	55	30189249	1	11382	10	11541	2	46448	1216825	24236	50	41163725	1	12192	10	12044	1	53177	1251716	26476	47	38303705	1	
34	ACETIC A	ACETIC A	5859	368425	4271	86	360681	1	2615	11	1656	2	5307	329945	3893	85	329502	1	2426	11	1467	1	5283	319471	3868	83	324115	1	
35	ACETYLCY	ACETYLCY	35634	1555543	14055	111	2464842	1	5865	15	8190	1	28352	1307055	10134	129	1864273	1	3650	18	6484	1	12363	1530784	5924	258	833930	2	
36	ACID JELL	ACETIC A	229	8237	136	61	23690	0	111	26	25	2	28	899	25	36	2465	0	null	null	null	null	null	null	null	null	null	null	nul
37	ACIPHEX	RABEPRA	404412	1.32E+08	88282	1494	18216153	7	44579	316	43703	21	372419	1.48E+08	76690	1926	17565458	8	40534	352	36156	21	284540	1.54E+08	68258	2250	15200860	10	
38	ACIPHEX	RABEPRA	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	nul

Database Constraints

1. I assigned INT, BIGINT, DECIMAL,VARCHAR type and size to each data.
2. I specified “NOT NULL AUTO_INCREMENT PRIMARY KEY” constraint for Brand_Code, Generic_Code, and Year_Code of each dimension table.
3. At the same time, I specified “FOREIGN KEY & REFERENCES” constraints for Brand_Code, Generic_Code, and Year_Code in the fact table.
4. Finally, I specified “CREATE INDEX” for columns 2 and 3 above. This is for fast query results.
5. I am not yet considering additional functions or procedures.

3. CODE to create a database

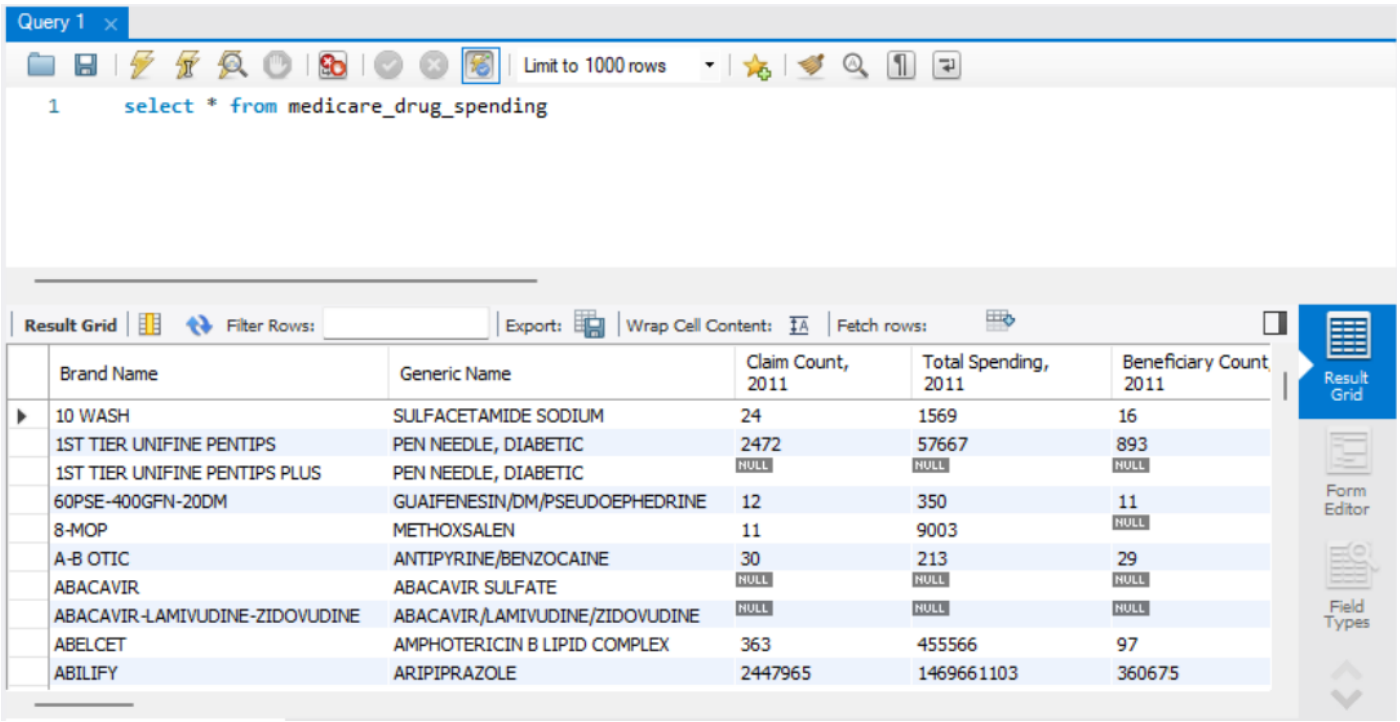
“CREATE DATABASE final_project”

Importing data to mySQL

Create database final_project

Table data import wizard -> importing

Select * from medicare_drug_spending



Brand Name	Generic Name	Claim Count, 2011	Total Spending, 2011	Beneficiary Count 2011
10 WASH	SULFACETAMIDE SODIUM	24	1569	16
1ST TIER UNIFINE PENTIPS	PEN NEEDLE, DIABETIC	2472	57667	893
1ST TIER UNIFINE PENTIPS PLUS	PEN NEEDLE, DIABETIC	NULL	NULL	NULL
60PSE-400GFN-20DM	GUAIFENESIN/DM/PSEUDOEPHEDRINE	12	350	11
8-MOP	METHOXSALEN	11	9003	NULL
A-B OTIC	ANTIPYRINE/BENZOCAINE	30	213	29
ABACAVIR	ABACAVIR SULFATE	NULL	NULL	NULL
ABACAVIR-LAMIVUDINE-ZIDOVUDINE	ABACAVIR/LAMIVUDINE/ZIDOVUDINE	NULL	NULL	NULL
ABELCET	AMPHOTERICIN B LIPID COMPLEX	363	455566	97
ABILIFY	ARIPIPRAZOLE	2447965	1469661103	360675

Create table

```
CREATE TABLE Brand (
        Brand_Code INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
    Brand_Name VARCHAR(160)
);

CREATE INDEX IDX_Brand_Name on Brand(Brand_Name);

CREATE TABLE Generic (
        Generic_Code INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
    Generic_Name VARCHAR(160)
);
CREATE INDEX IDX_Generic_Name on Generic(Generic_Name);

CREATE TABLE YearDim (
        Year_Code INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
    Year_Name VARCHAR(160)
);

CREATE TABLE medi_fact (
        Year_Code int,
    Brand_Code int,
    Generic_Code int,
    Claim_Count bigint(20) ,
        Total_Spending Decimal(65,2),
        Beneficiary_Count bigint(20) ,
        Total_Annual_Spending_per_User Decimal(65,2),
        Unit_Count bigint(20),
```

```
        Average_Cost_Per_Unit Decimal(65,2),
        Beneficiary_Count_No_LIS bigint(20) ,
        Beneficiary_Count_LIS bigint(20) ,
    PRIMARY KEY(Year_Code, Brand_Code, Generic_Code),
FOREIGN KEY (Brand_Code) REFERENCES Brand (Brand_Code),
FOREIGN KEY (Generic_Code) REFERENCES Generic (Generic_Code),
FOREIGN KEY (Year_Code) REFERENCES YearDim (Year_Code)
);

CREATE INDEX IDX_medifact_year on medi_fact(Year_Code);
CREATE INDEX IDX_medifact_brand on medi_fact(Brand_Code);
CREATE INDEX IDX_medifact_generic on medi_fact(Generic_Code);
```

Insert Data

```
insert into Brand (Brand_Name)
SELECT distinct `Brand Name`
FROM medicare
;
insert into Generic (Generic_Name)
SELECT distinct `Generic Name`
FROM medicare
;
insert into YearDim (Year_Name) values ('2011'),('2012'),('2013'),('2014'),('2015');

insert into medi_fact (
        Year_Code,
    Brand_Code,
```

```
    Generic_Code,
    Claim_Count ,
        Total_Spending,
        Beneficiary_Count,
        Total_Annual_Spending_per_User,
        Unit_Count,
        Average_Cost_Per_Unit,
        Beneficiary_Count_No_LIS,
        Beneficiary_Count_LIS
)
select (select year_code from yeardim where Year_Name = '2011'),
                (select Brand_Code from Brand where Brand_Name = t.`Brand Name`),
                (select Generic_Code from Generic where Generic_Name = t.`Generic Name`),
                `Claim Count, 2011`, `Total Spending, 2011`, `Beneficiary Count, 2011`, `Total Annual Spending Per User, 2011`,
                `Unit Count, 2011`, `Average Cost Per Unit (Weighted), 2011`, `Beneficiary Count No LIS, 2011`, `Beneficiary Count LIS, 2011`
from medicare t
union all
select (select year_code from yeardim where Year_Name = '2012'),
                (select Brand_Code from Brand where Brand_Name = t.`Brand Name`),
                (select Generic_Code from Generic where Generic_Name = t.`Generic Name`),
                `Claim Count, 2012`, `Total Spending, 2012`, `Beneficiary Count, 2012`, `Total Annual Spending Per User, 2012`,
                `Unit Count, 2012`, `Average Cost Per Unit (Weighted), 2012`, `Beneficiary Count No LIS, 2012`, `Beneficiary Count LIS, 2012`
from medicare t
union all
select (select year_code from yeardim where Year_Name = '2013'),
                (select Brand_Code from Brand where Brand_Name = t.`Brand Name`),
```

```
		(select Generic_Code from Generic where Generic_Name = t.`Generic Name`),
		`Claim Count, 2013`, `Total Spending, 2013`, `Beneficiary Count, 2013`, `Total Annual Spending Per User, 2013`,
		`Unit Count, 2013`, `Average Cost Per Unit (Weighted), 2013`, `Beneficiary Count No LIS, 2013`, `Beneficiary Count LIS, 2013`
from medicare t
union all
select (select year_code from yeardim where Year_Name = '2014'),
		(select Brand_Code from Brand where Brand_Name = t.`Brand Name`),
		(select Generic_Code from Generic where Generic_Name = t.`Generic Name`),
		`Claim Count, 2014`, `Total Spending, 2014`, `Beneficiary Count, 2014`, `Total Annual Spending Per User, 2014`,
		`Unit Count, 2014`, `Average Cost Per Unit (Weighted), 2014`, `Beneficiary Count No LIS, 2014`, `Beneficiary Count LIS, 2014`
from medicare t
union all
select (select year_code from yeardim where Year_Name = '2015'),
		(select Brand_Code from Brand where Brand_Name = t.`Brand Name`),
		(select Generic_Code from Generic where Generic_Name = t.`Generic Name`),
		`Claim Count, 2015`, `Total Spending, 2015`, `Beneficiary Count, 2015`, `Total Annual Spending Per User, 2015`,
		`Unit Count, 2015`, `Average Cost Per Unit (Weighted), 2015`, `Beneficiary Count No LIS, 2015`, `Beneficiary Count LIS, 2015`
from medicare t
;
select y.Year_Name, b.Brand_Name, g.Generic_Name, t.*
from medi_fact t
inner join brand b
    on b.Brand_Code = t.Brand_Code
inner join generic g
```

```
    on g.Generic_Code = t.Generic_Code
inner join yeardim y
    on y.Year_Code = t.Year_Code
```

4. Assessment Table (Individual Project)

Conceptual Schema: 10

Database: 10

Code: 10