DWBI PROJECT- Version 2.0



[**INFO7290 13752 Data Warehouse & Business Intelligence SEC 01 Fall 2018**](https://northeastern.blackboard.com/webapps/blackboard/execute/courseMain?course_id=_2565152_1)

**Table of Contents**

1. **Introduction……………………………………………………….………………………………………………………….….2**
2. **Overview of Each Data Set………………………………………………………………………….….………………….3**
   1. **HCIS……………………………………………………………………………………………….………………………3**
   2. **HPSA and Low Income……………………………….…………………………………………………………..4**
   3. **Medicare Provider Utilization and Payments Data………………………………………………….5**
   4. **Open Payments………………………………………………………………………………………………..…...6**
   5. **Zip Code Data (Income)………………………………………………………………………..………………..7**
   6. **State Fips……………………………………………………………………………………………………………….8**
   7. **County Fips……………………………………………………………………………………….…………….…….8**
3. **Summary Tables……………………………………………………………………………….……………………………….9**
4. **Visualization Plan………………………….………………………………………….…………………………………….24**

**Introduction**

Created data warehouse for the CMS Healthcare dataset which includes information related to physicians, teaching Hospitals and Payments related to insurance companies. linked all the different data set using a key that will help in joining them (zip-code). After that created different summary tables based on the insights retrieved after analyzing them. Visualized the summary tables which helped in decision making process.

Created Reports based on summary tables to derive insights, Patterns like problem of Medicare payments per state, Allotment of physicians based on the income groups, Spending done by the hospitals and physicians, finding patterns and ways to improve the revenue system of insurance companies. Identifying major areas which requires proper allotment of physicians based on the diseases and infection rate.

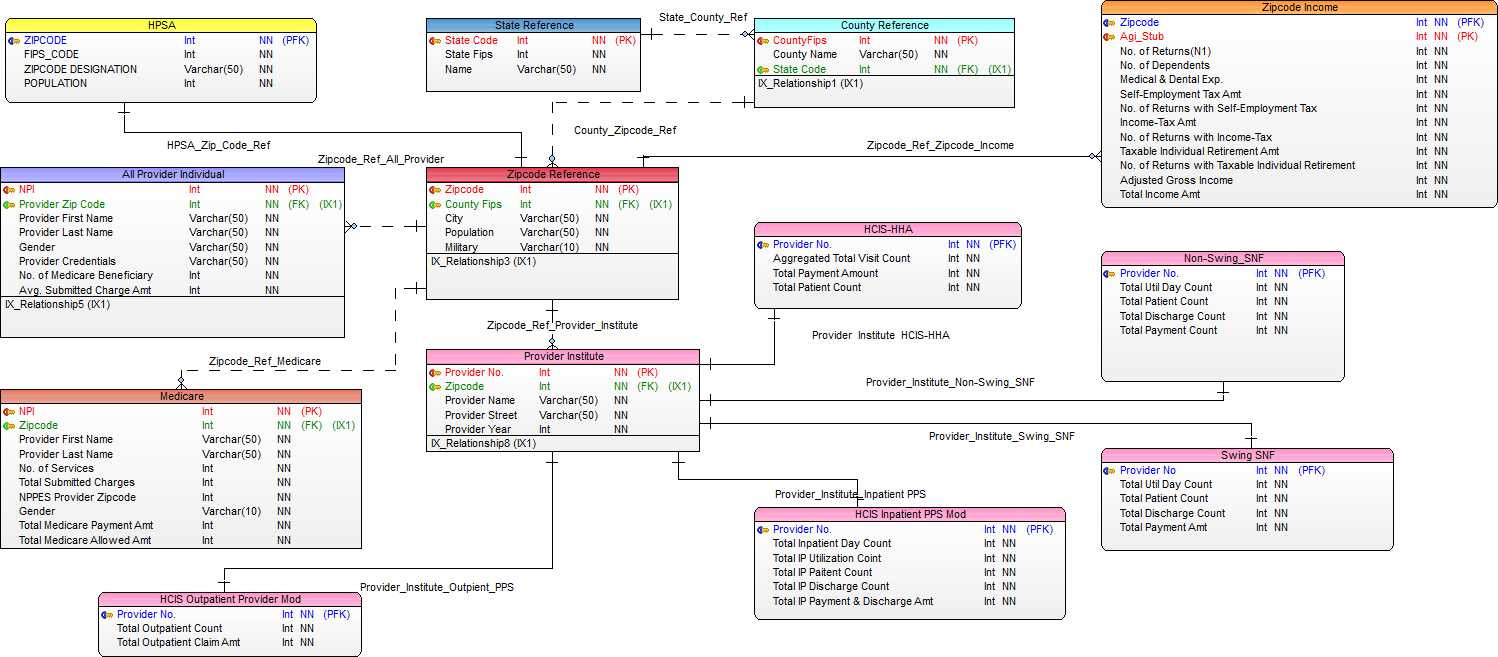
Developed summary and visualization dashboards related to spending for each episode Visits which informs us how and when are the skilled professionals appointed based on the paid HHRG (Home Health resource group- basis of payments for each episode)

Performed data analysis, star schema data modeling and design specific to data warehousing and business intelligence environment. Developed efficient SSIS packages for processing fact and dimension tables with complex transformation like type 1 and type 2 changes using slowly changing dimensions.

Used For-each loop to integrate data with 3 million rows into one staging area and performed analysis.

Optimized ETL performance to provide quickest response time possible. Created management reports with SSRS as well as SQL.

**Data Model:**



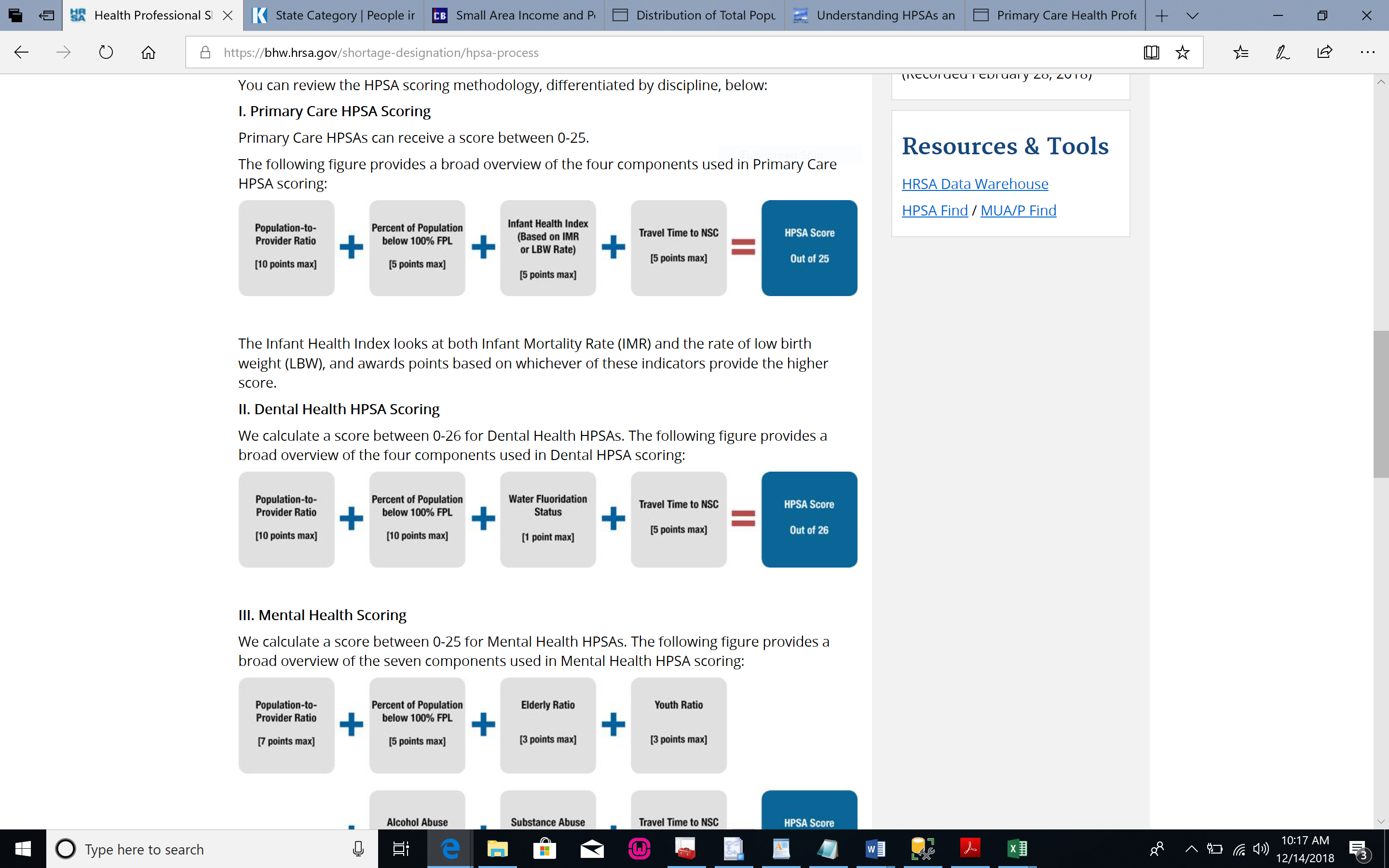
HPSA score based on 3 different discipline

1.Primary Care

2. Dental Care

3.Mental Health

HPSA Scoring Components:



**Overview of Each Data Set**

**HCIS**

HCIS stands for Health Care Information Systems. This data set contains data of 3 years from 2009 to 2011. HCIS contains data of inpatient and outpatient based on the type and State of the institutional provider. The types of provider included in the data set are Home Health Agency (HHA), Hospice, inpatient, Outpatient and Skilled Nursing Facility. HHA is an organization that gives home care services, like skilled nursing care, physical therapy, occupational therapy, speech therapy, and care by home health aides. Hospice care is provided for terminally ill patients and their families. It includes physical care as well as counseling services and is covered under Medicare Part A (hospital insurance). Skilled Nursing Facility is a facility that provides skilled nursing or rehabilitation services to help you recover after a hospital stay.

Prospective Payment System (PPS) is a predetermined base payment, which is adjusted for the health condition and care needs of the beneficiary as well as the geographic differences in wages for HHAs across the country.

The home health PPS is composed of six main features:

* Payment for the 60-day episode
* Case-Mix Adjustment
* Outlier Payments
* Adjustments for Low-Utilization Beneficiaries.
* Adjustments for beneficiaries with significant changes in condition.
* Adjustment for beneficiaries who change HHAs.

Main files of this data set are –

* Summary by HIPPS Medicare Payments - PPS
* Summary by HIPPS Medicare Payments per Average - PPS
* Summary by HIPPS Medicare Payments per episode average - PPS
* Summary by HIPPS for Visit Counts – PPS
* Summary by HIPPS for Visit Counts per average – PPS
* Summary by HIPPS for Visit Counts per episode average – PPS
* Summary by Outlier Claims for Outlier Payments
* Summary by Outlier Claims for Outlier Visits Payments - PPS
* Summary by Outlier Claims for Units of Service Totals – PPS
* Provider State Summary by Episode Type for Medicare Payments - PPS
* Provider State Summary by Episode Type for Patient Counts – PPS
* Provider State Summary by Episode Type for Visit Counts - PPS

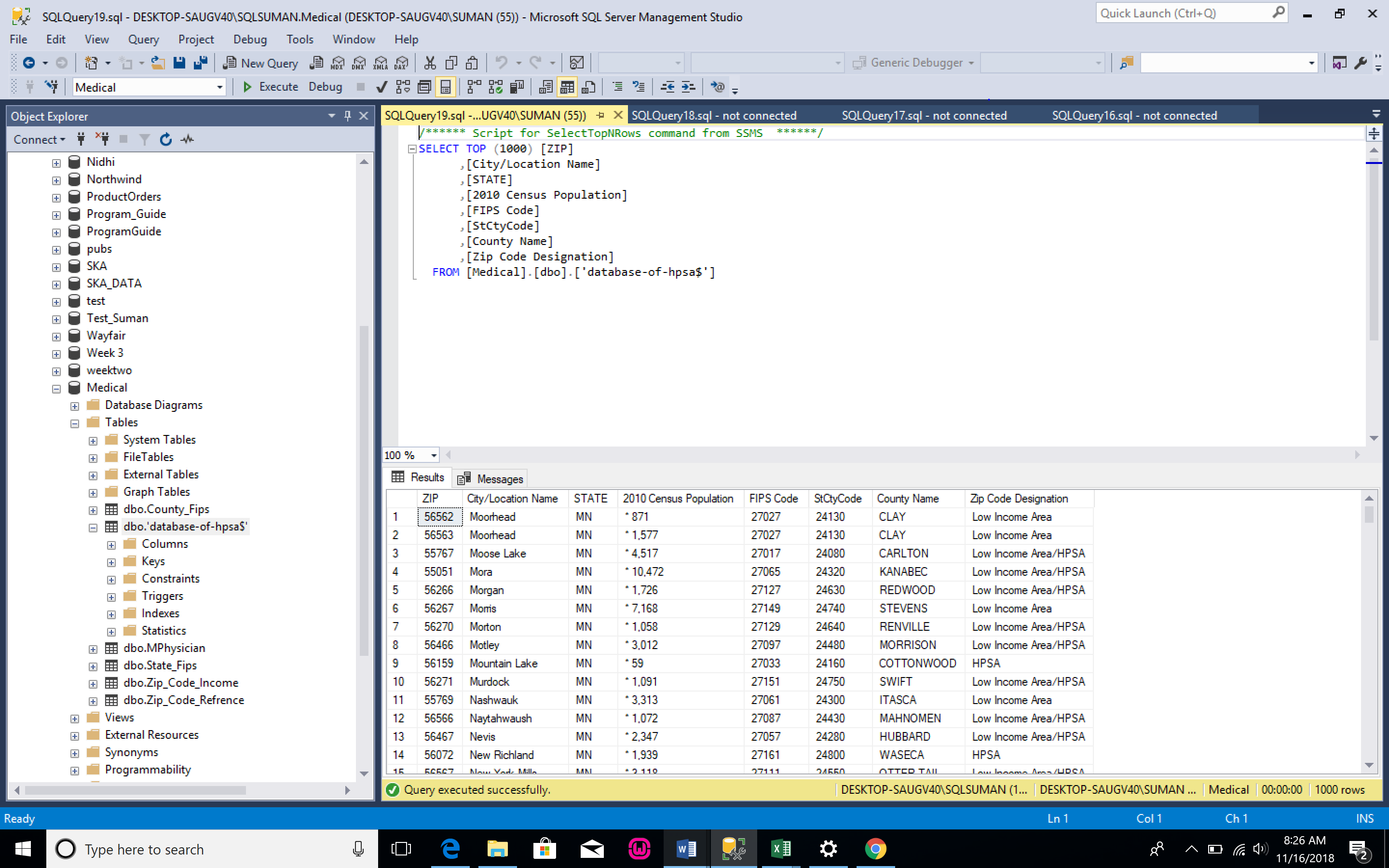
**HPSA & Low-Income**

Health Professional Shortage Area (HPSA)

HPSAs are geographic areas, or populations within geographic areas, that lack sufficient health care providers to meet the health care needs of the area or population.

* The Centers for Medicare & Medicaid Services (CMS) provides a 10 percent bonus payment when you furnish Medicare-covered services to beneficiaries in a geographic HPSA.
* If an area does not have a geographic primary care HPSA designation but has a geographic mental health HPSA designation, only psychiatrists who furnish services to Medicare beneficiaries in the designated area are eligible for the 10 percent bonus.
* If an area has both a primary care and a mental health geographic HPSA designation, only one HPSA bonus will be paid.

Here primary key id Zip code. It’s the one of the most important table as it is telling us about the Zip Code Designation which confirms the Zip Code Income group whether its Low-Income Area or HSPA. We can use this data to find the correlation between co-relation between different entities, like doctor and low-income group people. Correlation between zip codes for e.g. For $1-$25000 and $75000-$100000, same number of dependents but a smaller number of returns in second category.



**Medicare Provider Utilization and Payment Data**

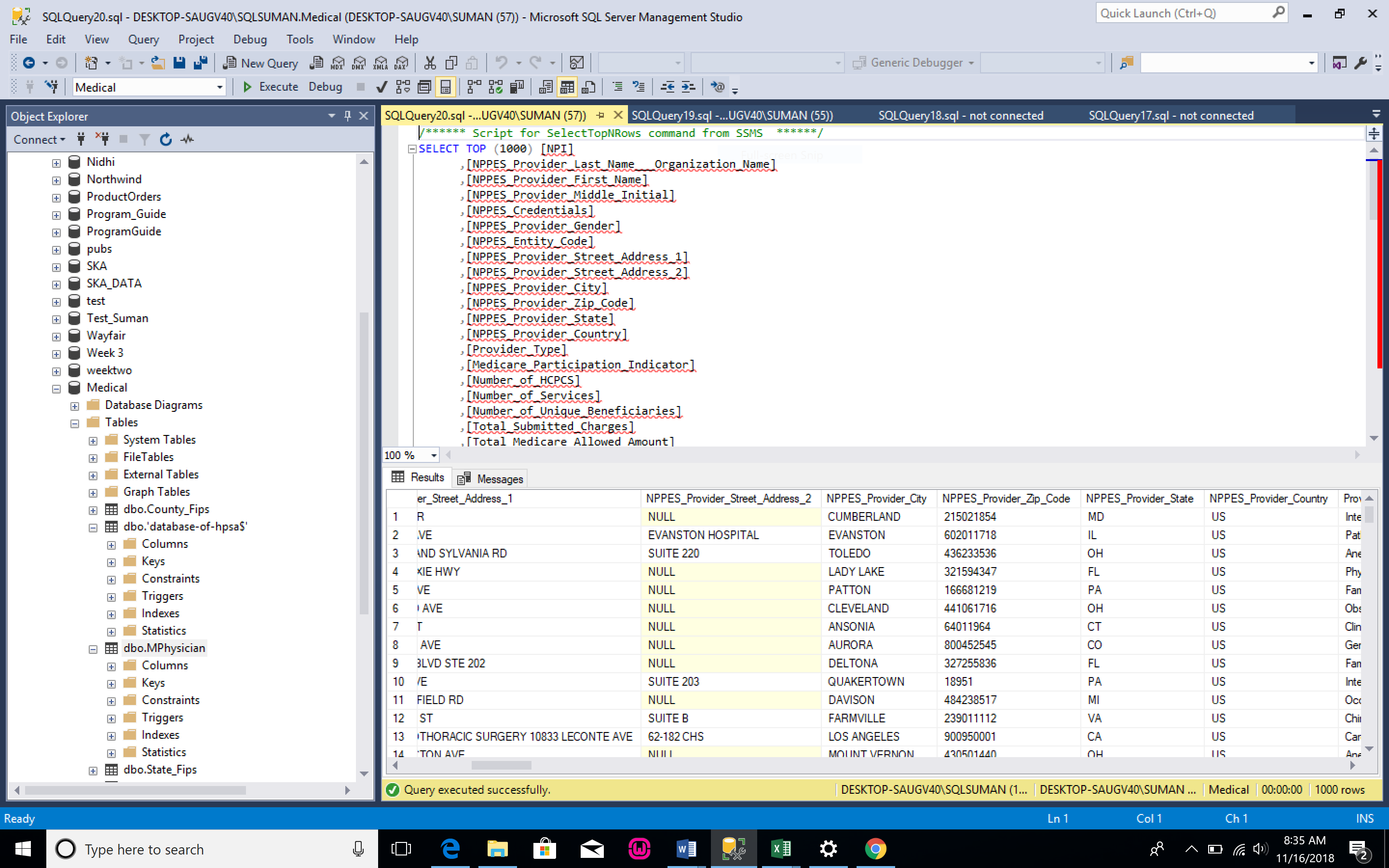
The Physician and Other Supplier Public Use File (Physician and Other Supplier PUF) provides information on services and procedures provided to Medicare beneficiaries by physicians and other healthcare professionals.

It contains information on utilization, payment (allowed amount and Medicare payment), and submitted charges organized by National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and place of service.

Medicare Physician and Other Supplier National Provider Identifier (NPI) Aggregate Report is a part of the detailed data set.

It contains information on utilization, payments (Medicare allowed amount, Medicare payment, standardized Medicare payment), and submitted charges organized by NPI.

In this table there is no direct Zip code is mentioned but we can extract the ZIP Code for the physician’s availability from ‘NPPES\_Provider\_Zip\_Code’ column as the 1St 5 digit represents the Zip code. Zip code will behave as the primary key to link with other data set.



**Open Payments**

* General Spend:

General Payments are defined as payments or other transfers made to a covered recipient (physician or teaching hospital) that are not made in connection with a research agreement or research protocol.

General Spend file provides the total value of general payments to a recipient recorded on a date.

* Research Spend:

Research Payments are defined as payments or other transfers made in connection with a research agreement or research protocol.

Research Spend file provides the total value of a payment made for research purposes to a recipient recorded on a date. Information is also provided for up to five physician principal investigators associated with the payment.

* Ownership Spend:

Ownership and Investment Interest Information is defined as information on the value of ownership or investment interests that a physician or an immediate family member of a physician held in an applicable manufacturer or applicable group purchasing organization (GPO).

Records for all three payment categories (general, research, physician ownership) include a Change Type which can be explained as follows:

|  |  |
| --- | --- |
| **Change Type Values** | **Meaning** |
| NEW | The payment record was submitted during the most recent submission window and is being published for the first time. |
| ADD | The payment record had been submitted prior to the most recent submission window but had been ineligible for publication until the current publication. The record is being published for the first time |
| CHANGED | The payment record has been published in a previous publication and has been modified since its last publication, which may include an update to its dispute status. |
| UNCHANGED | The payment record was published in the previous publication and is being republished without change in the current publication. |

**Zip Code Data (Income)**

This data set includes returns filed during the 12-month period, January 1, 2017 to December 31, 2017.

SOI did not attempt to correct any ZIP codes on the returns; however, it did take the following precautions to avoid disclosing information about specific taxpayers:

* ZIP codes with less than 100 returns and those identified as a single building or nonresidential ZIP code were categorized as “other” (99999).

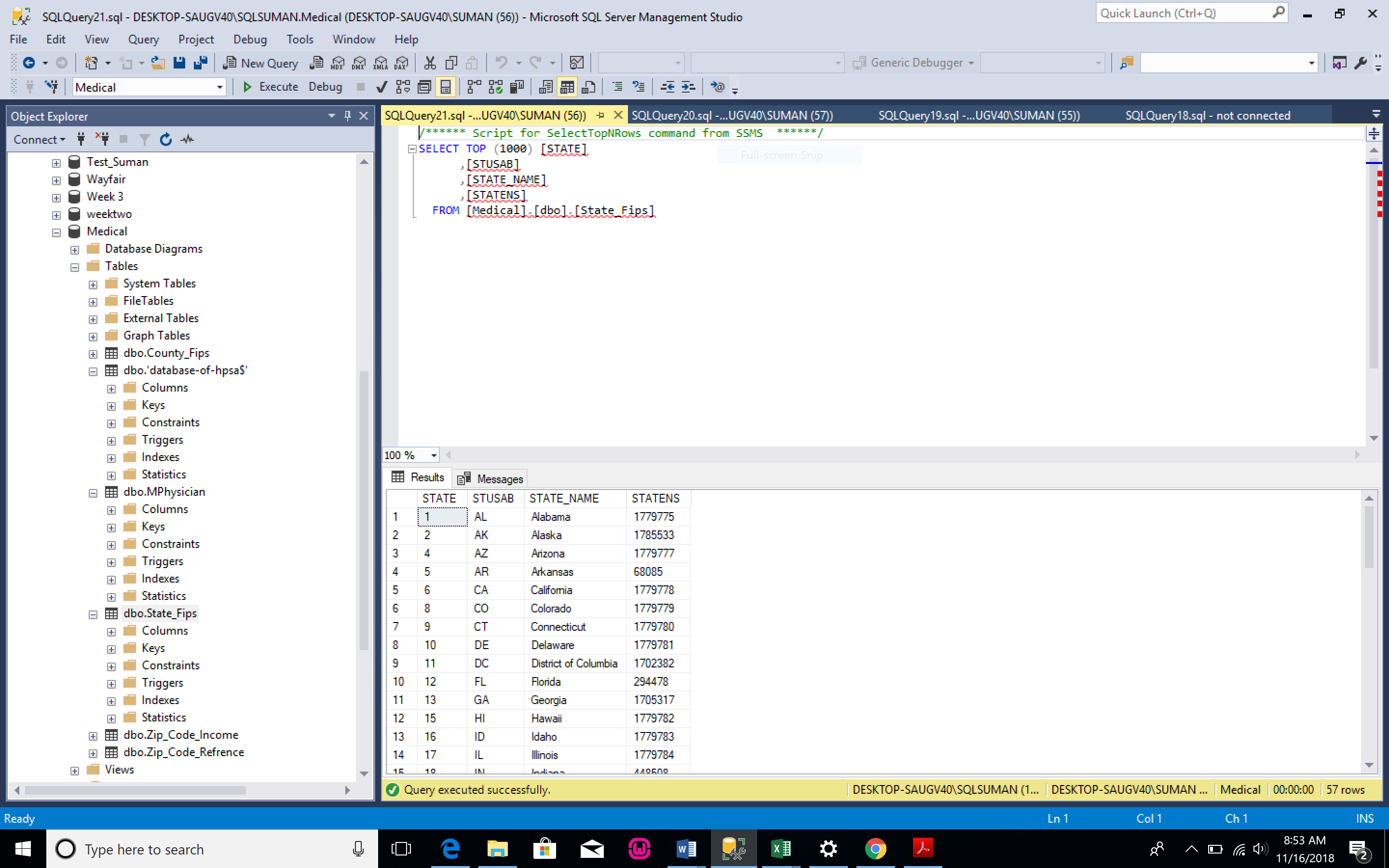
* Income and tax items with less than 20 returns for a particular AGI class were combined with another AGI class within the same ZIP Code. Collapsed AGI classes are identified with a double asterisk (\*\*).
* All number of returns variables have been rounded to the nearest 10.
* Excluded from the data are items with less than 20 returns within a ZIP code.
* Excluded from the data are tax returns with a negative adjusted gross income.

Excluded are tax returns representing a specified percentage of the total of any particular cell. For example, if one return represented 75 percent of the value of a given cell, the return was suppressed from the tabulation. The actual threshold percentage used cannot be released.

**State Fips**

This file contains pipe delimited records for each state. The records are of the format:

FIPS State Code | Official United States Postal Service (USPS) Code | Name | Geographic Names Information System Identifier



**County Fips**

This file contains the data in the below mentioned format. It gives the information about the counties of a state with its FIPS (Federal Information Processing Standards)

|  |  |  |
| --- | --- | --- |
| **Columns** | **Description** | **Sample** |
| STATE | State Postal Code | FL |
| STATEFP | State FIPS Code | 12 |
| COUNTYFP | County FIPS Code | 011 |
| COUNTYNAME | County Name and Legal/Statistical Area Description | Broward County |
| CLASSFP | FIPS Class Code | H1 |

**Summary Tables**

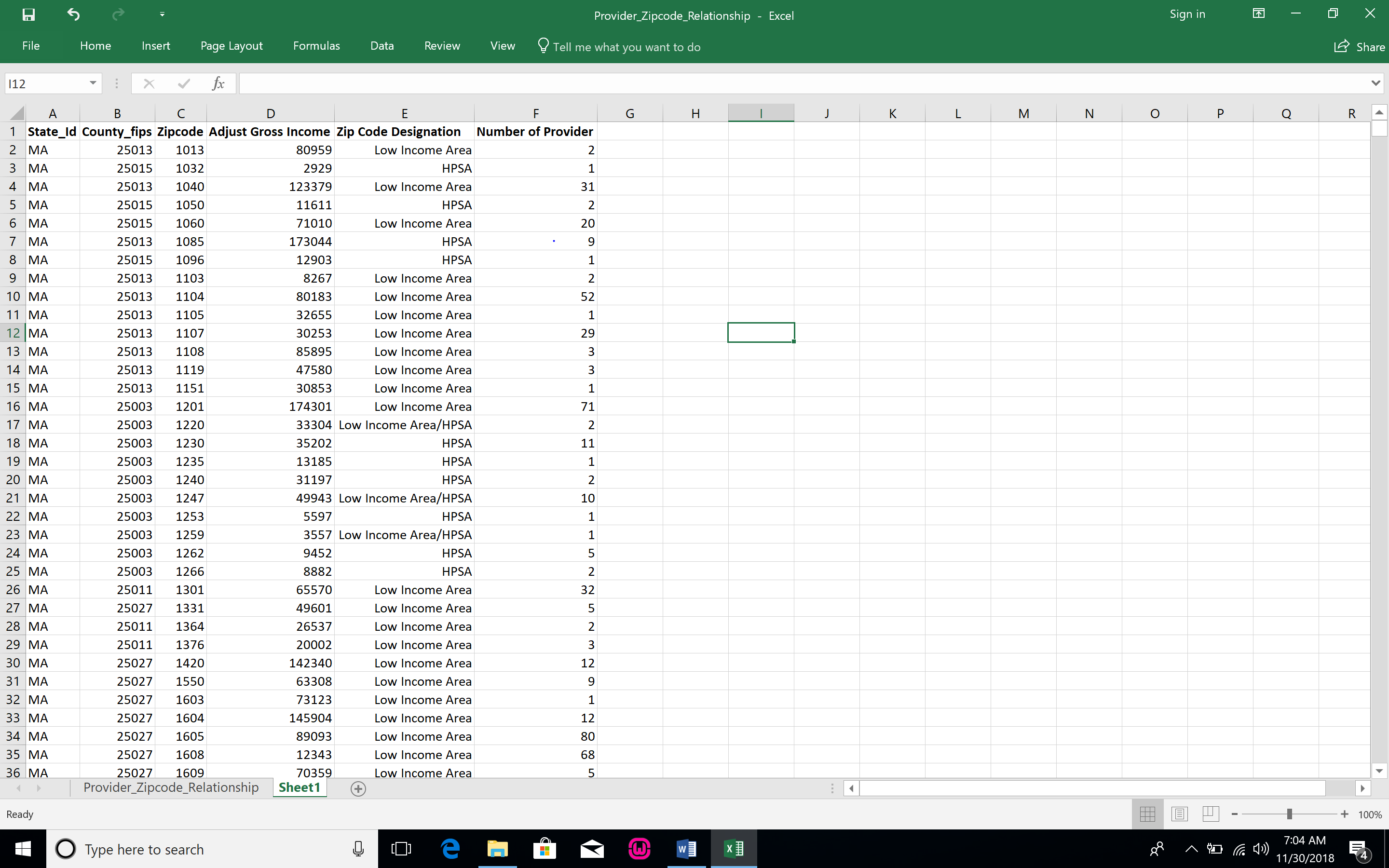
**Provider\_Zipcode\_Gross Income\_Relationship Data set**

This symmary table is for the showing that how may provider are present for each Zipcode plus whether that zipcode is considered as Low Income Area/HPSA/Low Income Area and HPSA.It also so the Average Adjust Gross Income for corresponding Zip code and County\_FIps and State\_ID for corresponding Zipcode. We can use report to provide multiple insights to the business.

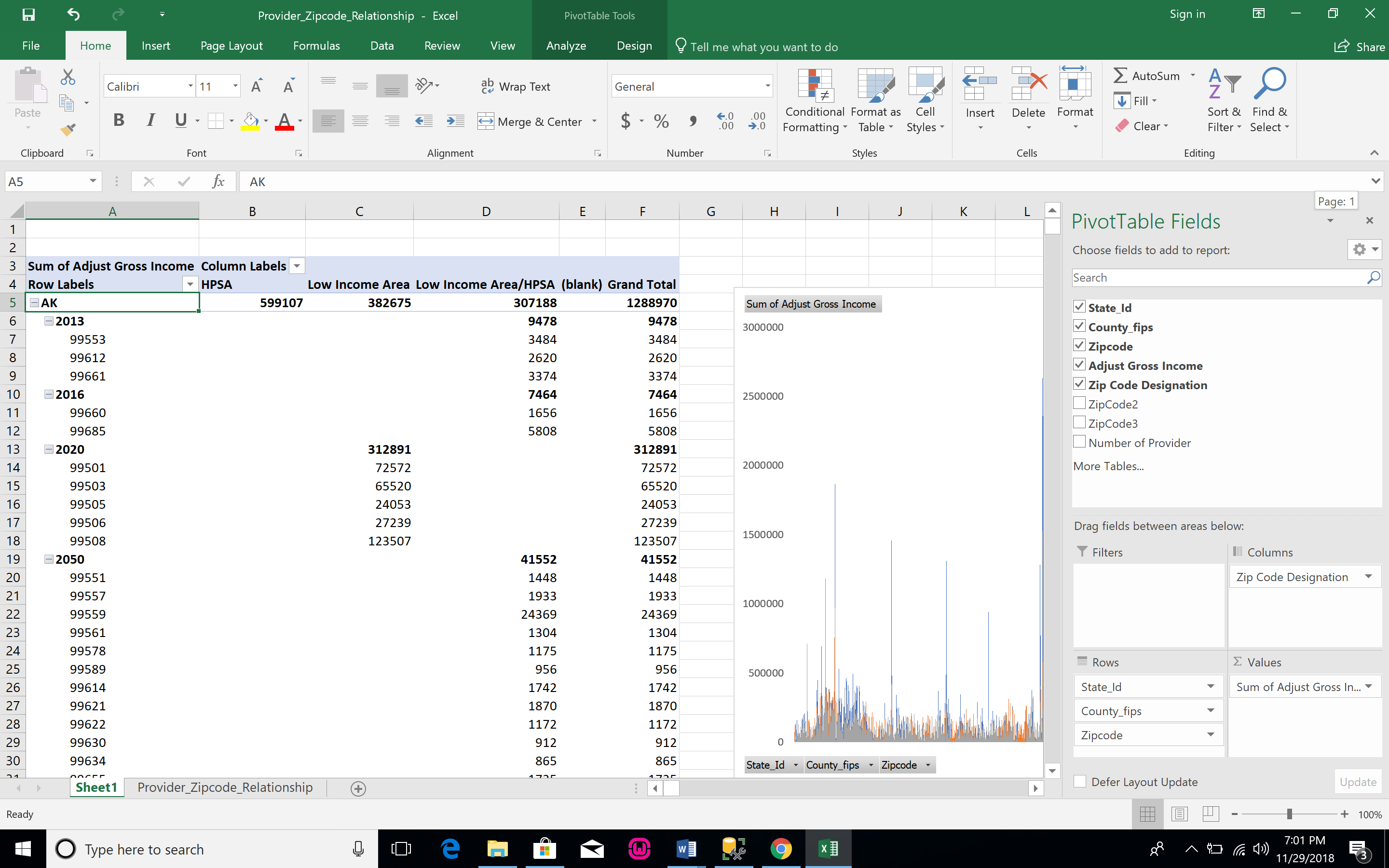
To create this report we have join for different data set those are HPSA dataset, Medicare Physician,ZipCode\_Income, Zipcode\_Reference. We have use ZipCode as a primary key for all file except Medicare Physician. The primary key for Medicare Phyician File is NPI. In this table Zip code is not given directly we have to extract it from ‘NPPES\_Provider\_Zip\_Code’. The 1st 5 digit is Zip code.

First we calculated the Adjust Gross Income which is ‘average of Gross Income’ paid by that Zip code, then join this file with ‘HPSA dataset’ to get the ‘ZIP Code Designation’ for corresponding Zip code; further join the new table with ‘Mediare Phyician’ table to get the number of provider for that particular zip code and at last we have joined the above new table with ZipCode refrence table to extarct State\_ID and County\_Fips.Please find the below summary table for the same

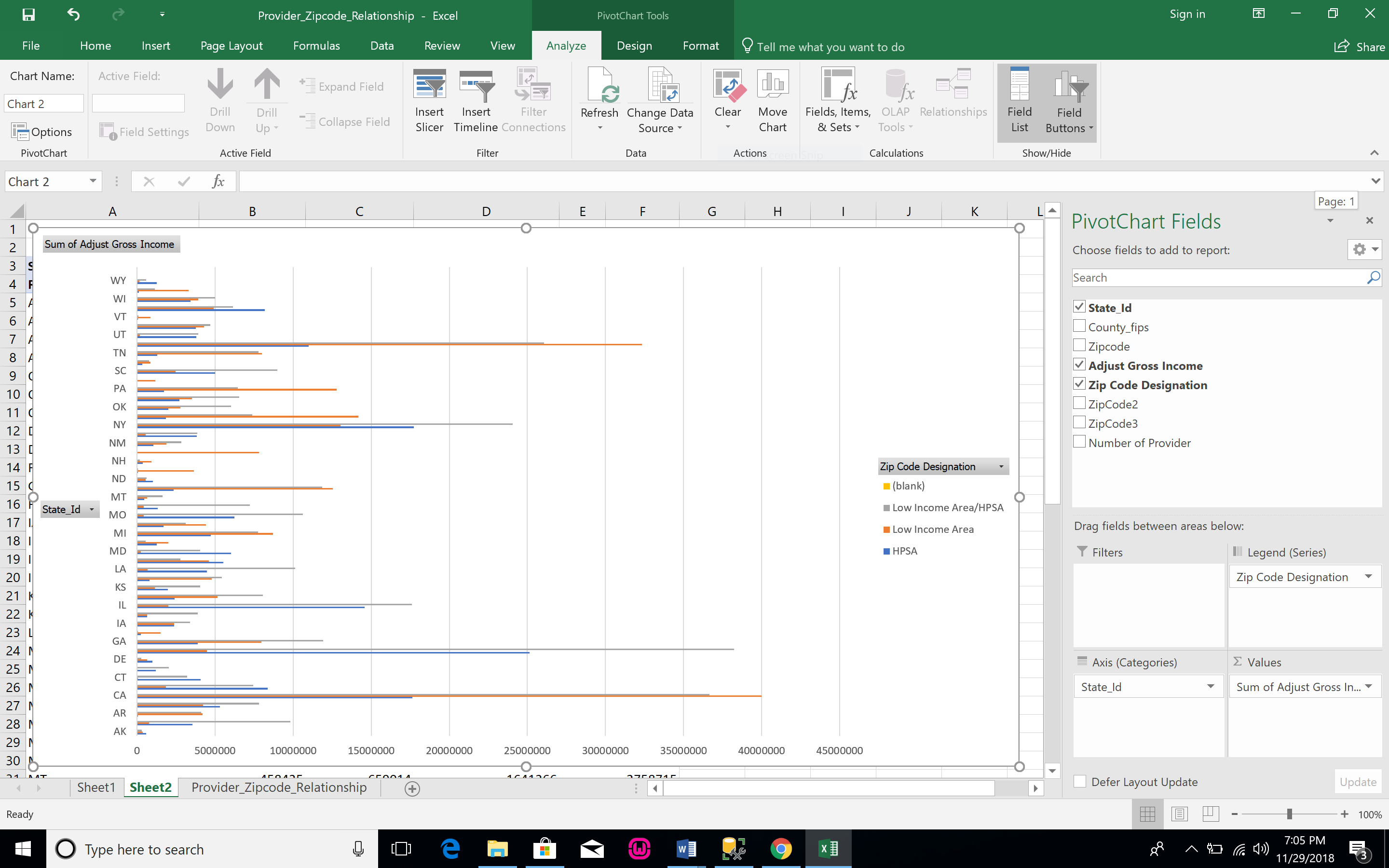
Note: We have considered full data for this and we have done some cleaning of data to provide the better report



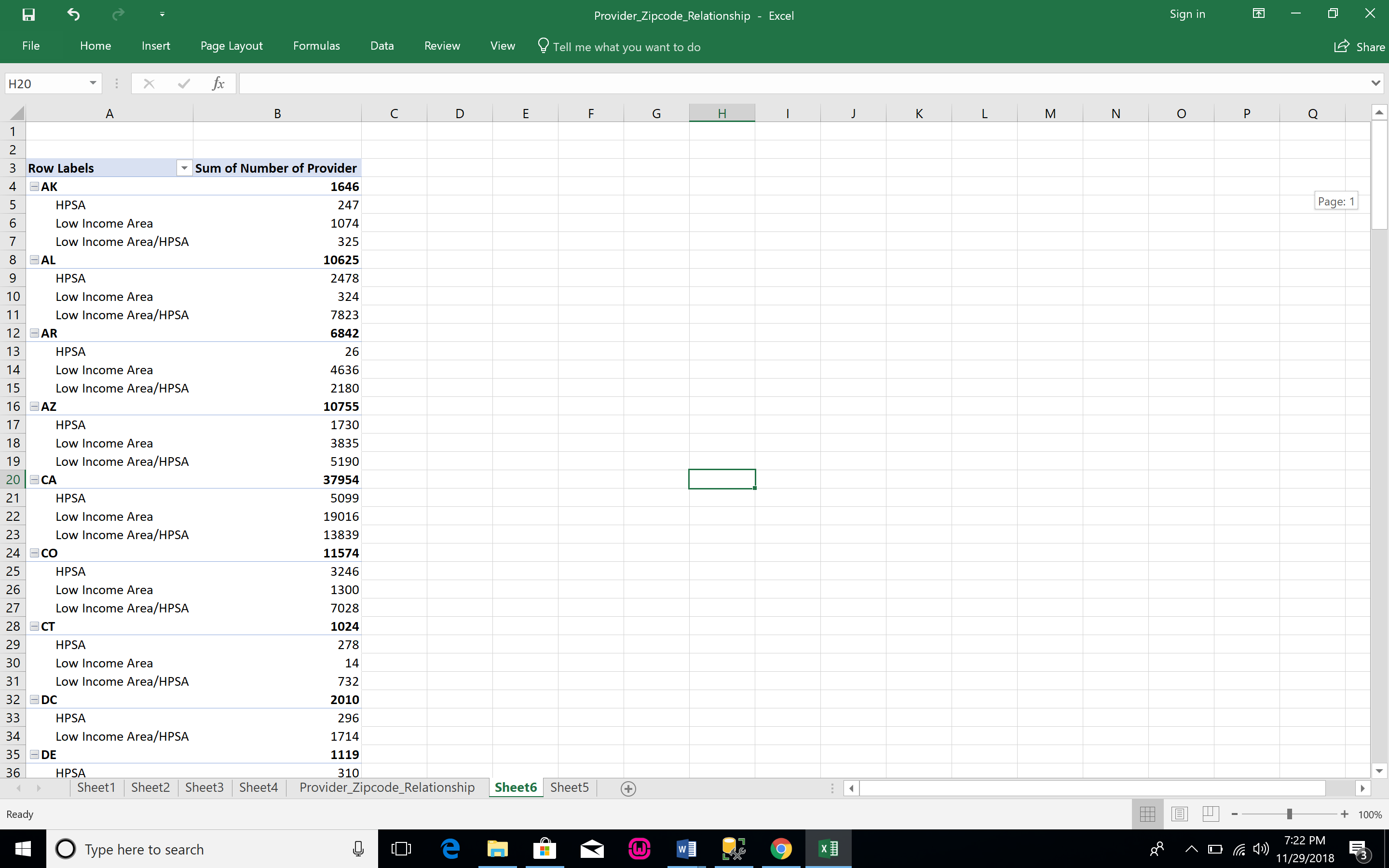
We have used Pivot table to further break the report for visulazition purpose to show Total Adjust Gross Income for ZipCode, County,State prospective plus it also show Adjust Gross Income for different Zip code designation.Please find the below screenshot for the same.



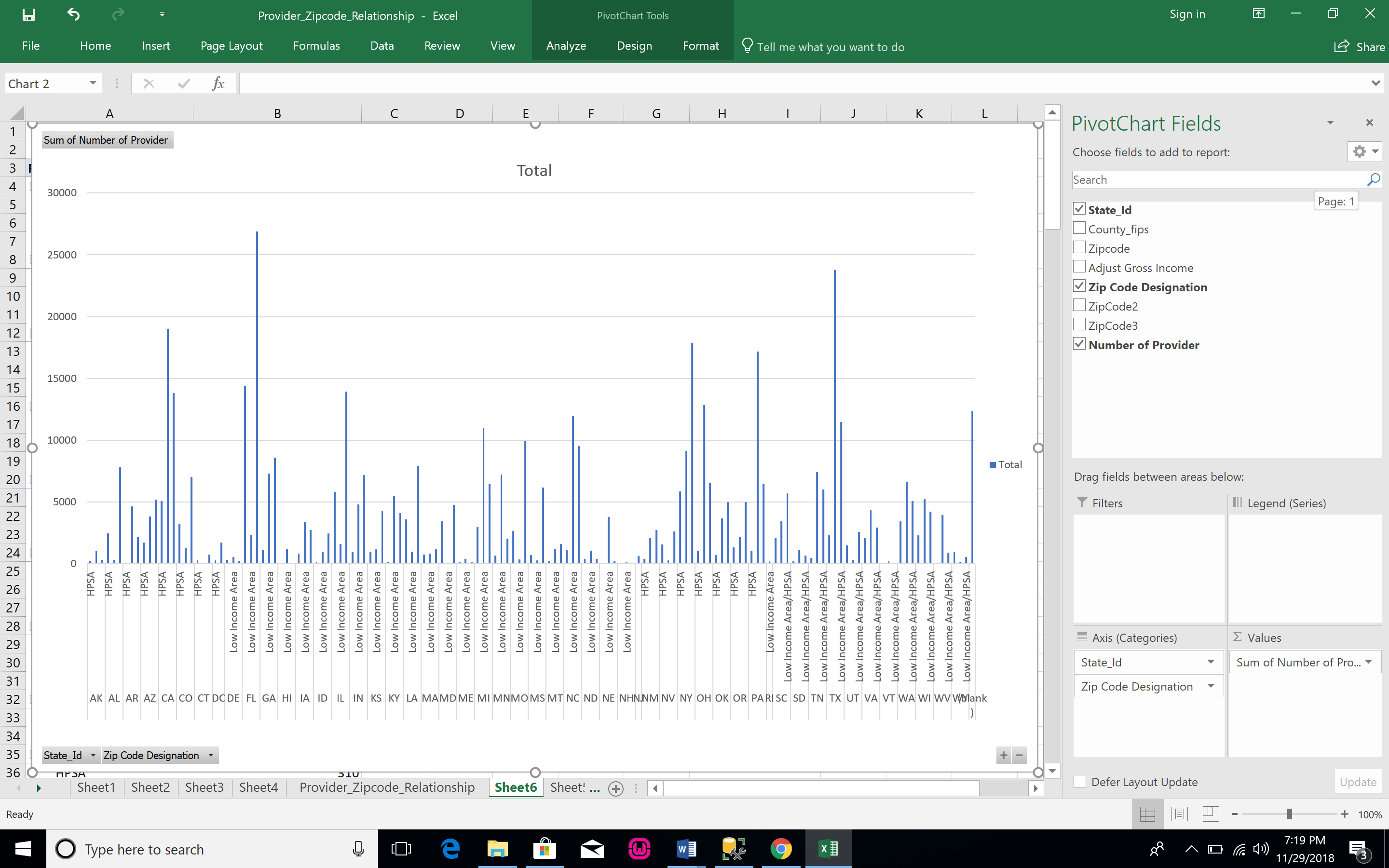
The Below chart shows the Sum of Adjust Gross Income for partiulcar state plus the colured line shows the Zip code designation Adjust Gross Income for particular state. We can use this chart to analysis and provide a solutions that how low income affects the HPSA. Plus we can also analysis is HPSA is totally depended on income or we used consider other factors as well to calculate HPSA.



We Further Break the Summary table using Pivot Table to find the number of provider in any Area(State, County, ZIpcode) plus how many of them belongs to which kind of ZipCode Designation(HPSA/Low Income Area/ Low Income Area and HPSA).



The below charts shows the number of Provider for the Particular state and breakage of number of provider on Zip code Designation basis.We can use this table and chart to analysis how many for personal is required in a particular area to remove it for HPSA list.



**Summary table for HCIS**

HHRG (Home Health Resource Group) are the basis of payment for each episode. It is a payment code which identifies 1 out of 80 possible home health episode payments under the PPS. A beneficiary is assigned to an HHRG based on three measures: clinical severity, functional status, and service utilization.

The below is a summary table for the outlier Visits which informs us how and when are the skilled professionals appointed based on the paid HHRG.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Paid HHRG** | **Skilled Nursing** | **Physical Therapy** | **Occupational Therapy** | **Speech Pathology** | **Medical Social Services** | **Home Health Aide** | **Total** |
| C1F1S1 | 86773 | 4083 | 1545 | 508 | 1332 | 39912 | 134153 |
| C1F1S2 | 4769 | 4962 | 3043 | 1227 | 298 | 4794 | 19093 |
| C1F1S3 | 8188 | 4581 | 2469 | 925 | 229 | 5727 | 22119 |
| C1F2S1 | 34145 | 7895 | 1965 | 550 | 1711 | 97899 | 144165 |
| C1F2S2 | 4063 | 7992 | 4382 | 2236 | 430 | 10225 | 29328 |
| C1F2S3 | 7540 | 11590 | 6145 | 2936 | 719 | 16844 | 45774 |
| C1F3S1 | 917238 | 3780 | 1394 | 270 | 440 | 36050 | 959172 |
| C1F3S2 | 54377 | 10676 | 6446 | 2433 | 426 | 8847 | 83205 |
| C1F3S3 | 9594 | 36607 | 21989 | 8951 | 1041 | 34890 | 113072 |
| C2F1S1 | 1518067 | 30376 | 3996 | 368 | 2997 | 81909 | 1637713 |
| C2F1S2 | 92130 | 1483 | 274 | 31 | 123 | 2105 | 96146 |
| C2F1S3 | 140794 | 325956 | 89451 | 24849 | 33230 | 1482421 | 2096701 |
| C2F3S1 | 40763 | 16911 | 2539 | 376 | 4090 | 199897 | 264576 |
| C2F3S2 | 5603 | 329429 | 90275 | 24937 | 33423 | 1490738 | 1974405 |
| C2F3S3 | 14431 | 10253 | 1628 | 151 | 672 | 19827 | 46962 |
| C3F1S1 | 33113 | 4972 | 984 | 182 | 998 | 88428 | 128677 |
| C3F1S2 | 133922 | 2291 | 375 | 49 | 149 | 5112 | 141898 |
| C3F1S3 | 4975 | 3405 | 552 | 104 | 199 | 7633 | 16868 |
| C3F2S1 | 821322 | 9915 | 1771 | 186 | 2838 | 47038 | 883070 |
| C3F2S2 | 41113 | 3730 | 567 | 80 | 113 | 1785 | 47388 |
| C3F2S3 | 7915 | 4945 | 730 | 102 | 189 | 2576 | 16457 |
| C3F3S1 | 496679 | 45072 | 6581 | 762 | 6628 | 279053 | 834775 |
| C3F3S2 | 17650 | 21917 | 3340 | 272 | 655 | 20022 | 63856 |
| C3F3S3 | 109413 | 25083 | 5138 | 434 | 1044 | 26418 | 167530 |
| C2F2S1 | 456554 | 20526 | 4544 | 710 | 4146 | 387048 | 873528 |
| C2F2S2 | 7002215 | 9396 | 2045 | 274 | 383 | 15920 | 7030233 |
| C2F2S3 | 7013837 | 13211 | 3147 | 476 | 629 | 26408 | 7057708 |
| Unkn | 0 | 2468 | 1272 | 446 | 513 | 27920 | 32619 |

Below is the summary table for the Outlier claims which informs us about the total Medicare payments for outlier claims based on Paid HHRG.

|  |  |  |  |
| --- | --- | --- | --- |
| **Paid HHRG** | **All Prorated Non-RAP claims** | **All Outliers Non\_RAP claims** | **Total Medicare Payments** |
| C1F1S1 | 259539 | 3151 | $7,316,100.00 |
| C1F1S2 | 20377 | 107 | $1,769,225.00 |
| C1F1S3 | 33732 | 166 | $1,667,020.00 |
| C1F2S1 | 420075 | 4255 | $11,651,093.00 |
| C1F2S2 | 53949 | 290 | $3,243,108.00 |
| C1F2S3 | 92133 | 434 | $4,821,463.00 |
| C1F3S1 | 96922 | 1254 | $4,909,874.00 |
| C1F3S2 | 13021 | 73 | $5,639,864.00 |
| C1F3S3 | 20023 | 143 | $20,095,073.00 |
| C2F1S1 | 801160 | 11385 | $32,568,489.00 |
| C2F1S2 | 75889 | 398 | $825,651.00 |
| C2F1S3 | 151312 | 771 | $626,787,421.00 |
| C2F3S1 | 326629 | 5588 | $80,287,095.00 |
| C2F3S2 | 40131 | 245 | $628,923,797.00 |
| C2F3S3 | 52879 | 384 | $6,344,471.00 |
| C3F1S1 | 411950 | 17801 | $23,528,757.00 |
| C3F1S2 | 20238 | 416 | $1,260,950.00 |
| C3F1S3 | 27478 | 598 | $2,086,743.00 |
| C3F2S1 | 1111785 | 37730 | $70,456,505.00 |
| C3F2S2 | 130670 | 2135 | $2,277,731.00 |
| C3F2S3 | 173469 | 2847 | $3,236,850.00 |
| C3F3S1 | 652709 | 24039 | $181,113,798.00 |
| C3F3S2 | 62227 | 1010 | $14,232,511.00 |
| C3F3S3 | 83639 | 1589 | $17,324,442.00 |
| C2F2S1 | 581268 | 14010 | $107,797,979.00 |
| C2F2S2 | 94614 | 483 | $6,495,759.00 |
| C2F2S3 | 80194 | 740 | $9,477,124.00 |

From the graph it is evident that the claims referring to C2F1S3 and C2F3S2 uses the majority of Medicare Payments.

**HPSA Dataset Summary Table**

This summary table gives us the information about the state and its total population which was given in terms of city and zip code. Also, this table will inform us about the state that has highest population which is evident in the graph below.

The summary table will also give us the counts of Low Income/HPSA, HPSA and Low Income belonging to a specific state. We have visualized the data for the highly populated states.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STATE** | **Population** | **Count\_Lo/HPSA** | **Count\_HPSA** | **Count\_LowIcome** | **Count\_Total** |
| AK | 309077 | 126 | 25 | 11 | 162 |
| AL | 3953366 | 456 | 80 | 44 | 580 |
| AR | 2519673 | 319 | 7 | 214 | 540 |
| AZ | 4387176 | 145 | 65 | 66 | 276 |
| CA | 20922781 | 545 | 253 | 308 | 1106 |
| CO | 2783656 | 195 | 166 | 35 | 396 |
| CT | 1092983 | 39 | 40 | 4 | 83 |
| DC | 468209 | 13 | 9 | 0 | 22 |
| DE | 319543 | 3 | 6 | 16 | 25 |
| FL | 13464368 | 513 | 229 | 49 | 791 |
| GA | 6385690 | 431 | 69 | 111 | 611 |
| HI | 434856 | 8 | 8 | 25 | 41 |
| IA | 1902690 | 297 | 381 | 90 | 768 |
| ID | 1296565 | 199 | 31 | 12 | 242 |
| IL | 6941203 | 498 | 612 | 29 | 1139 |
| IN | 4175772 | 241 | 170 | 95 | 506 |
| KS | 1861566 | 374 | 194 | 51 | 619 |
| KY | 3287494 | 381 | 28 | 249 | 658 |
| LA | 3964032 | 350 | 77 | 38 | 465 |
| MA | 2001094 | 25 | 68 | 60 | 153 |
| MD | 1813372 | 65 | 111 | 12 | 188 |
| ME | 897083 | 93 | 20 | 178 | 291 |
| MI | 5640325 | 371 | 117 | 179 | 667 |
| MN | 2065587 | 221 | 212 | 122 | 555 |
| MO | 4145104 | 732 | 186 | 12 | 930 |
| MS | 2794774 | 343 | 21 | 16 | 380 |
| MT | 685464 | 188 | 55 | 61 | 304 |
| NC | 7236269 | 365 | 57 | 206 | 628 |
| ND | 396655 | 138 | 220 | 8 | 366 |
| NE | 996501 | 29 | 15 | 295 | 339 |
| NH | 335184 | 16 | 38 | 26 | 80 |
| NJ | 2222387 | 0 | 0 | 82 | 82 |
| NM | 1658466 | 206 | 22 | 45 | 273 |
| NV | 1420146 | 61 | 58 | 8 | 127 |
| NY | 10493695 | 551 | 361 | 183 | 1095 |
| OH | 6292015 | 253 | 57 | 319 | 629 |
| OK | 3050438 | 434 | 75 | 59 | 568 |
| OR | 2855573 | 236 | 76 | 56 | 368 |
| PA | 5748946 | 388 | 91 | 404 | 883 |
| PR | 3725789 | 41 | 0 | 78 | 119 |
| RI | 317811 | 0 | 0 | 14 | 14 |
| SC | 3774042 | 290 | 58 | 28 | 376 |
| SD | 459627 | 181 | 62 | 52 | 295 |
| TN | 4575280 | 311 | 32 | 182 | 525 |
| TX | 16965596 | 787 | 229 | 461 | 1477 |
| UT | 1622663 | 110 | 45 | 23 | 178 |
| VA | 2950661 | 290 | 134 | 152 | 576 |
| VT | 264127 | 15 | 9 | 85 | 109 |
| WA | 3627444 | 229 | 150 | 51 | 430 |
| WI | 2665664 | 231 | 143 | 82 | 456 |
| WV | 1577782 | 207 | 11 | 322 | 540 |
| WY | 354785 | 40 | 72 | 10 | 122 |

The graph shows that CA and TX are the highly populated areas.

CA has 49% - Low Income/HPSA, 23% - HPSA and 28% - Low Income counts.

TX has 53% - Low Income/HPSA, 16% - HPSA and 31% - Low Income counts.

**Medicare Data Set Summary Tables**

CMS has created two summary tables:

1) Aggregated information by physician or another supplier

2) Aggregated information by State and HCPCS code.   A detailed methodology document can be found in the Downloads section below which contains important information regarding the limitations of data usage.

Combining of datasets:

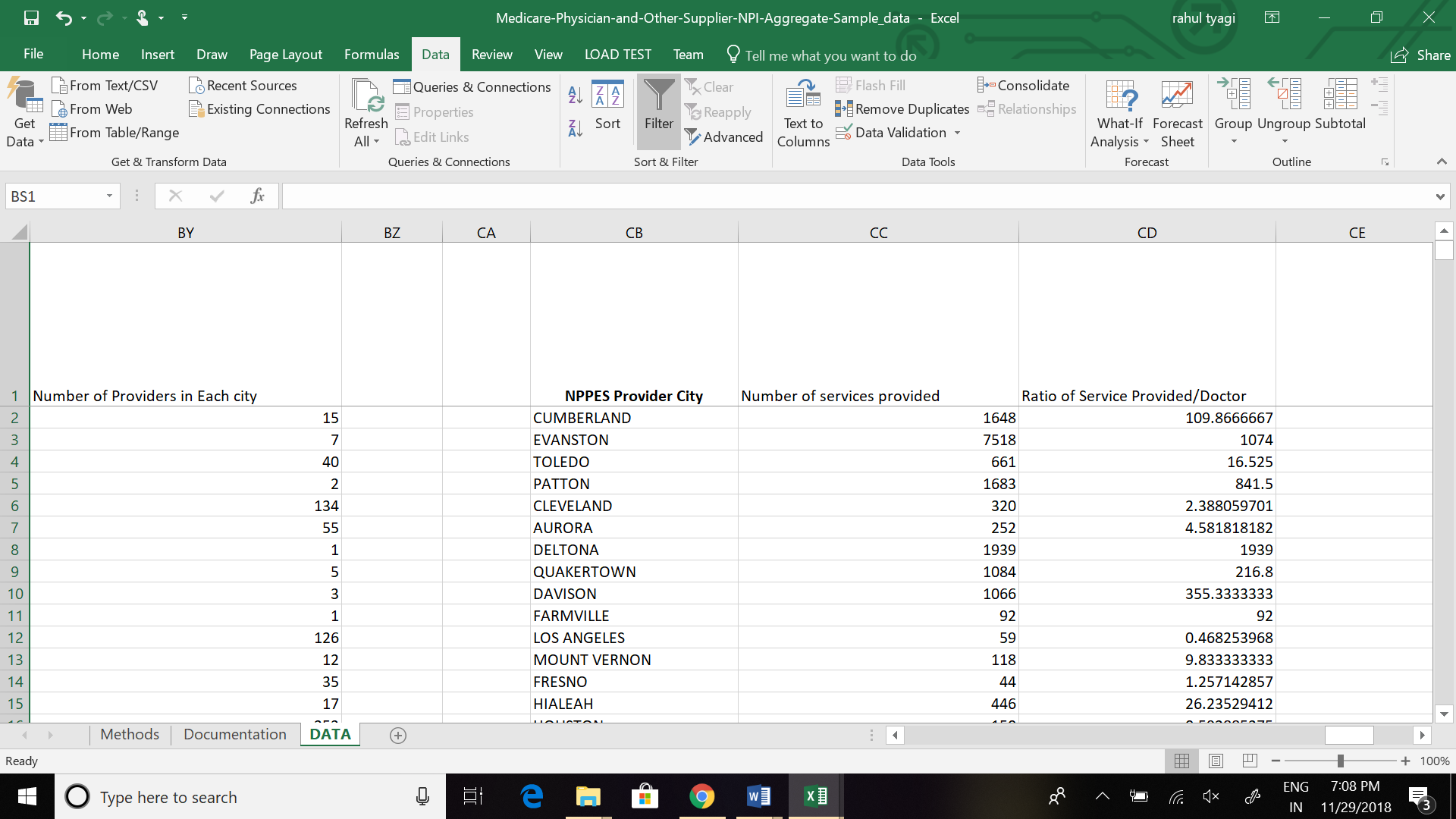
For the dataset of Aggregated information by State and HCPCS code I can see that HCPCS column is same, so we can link the two tables by using HCPCS (Health Care Procedure Coding System)

In the dataset aggregated information by physician or other supplier I can see a unique column NPPES provider Zip code which can be used as primary key to connect with other databases.

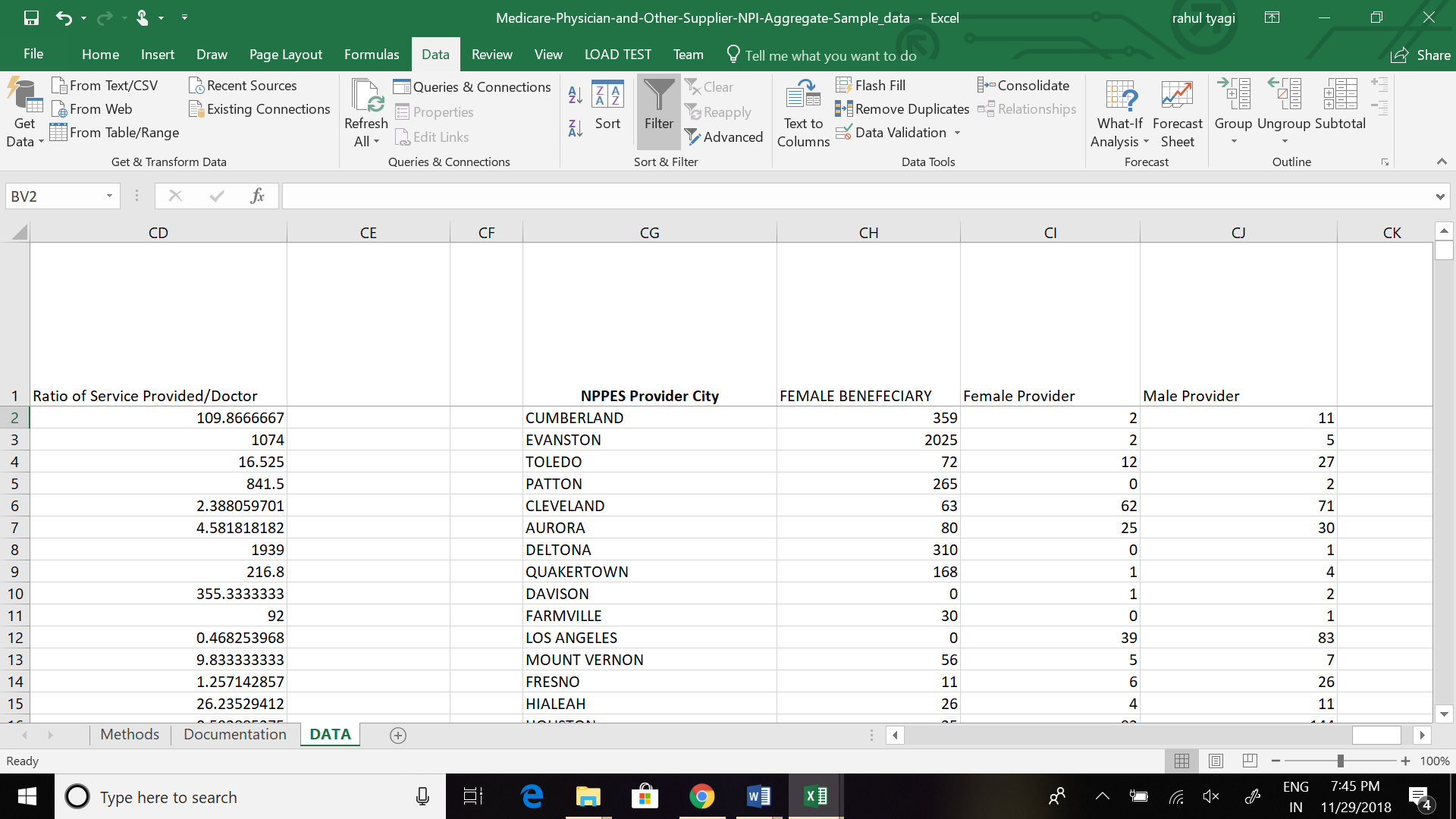
Screen-Shot of Dataset:



We have calculated number of providers which are providing services in cities and sum of all the services provided in a particular city. After that we calculated the Ratio of services provided/Doctor which will further help in deciding the assigning of doctors to cities based on the load of services. We can see from the table in some areas there are 92 patients/ doctor and in some areas 2 or 3 patients per doctor. In some areas there are more doctors and in some other areas there are less doctors.

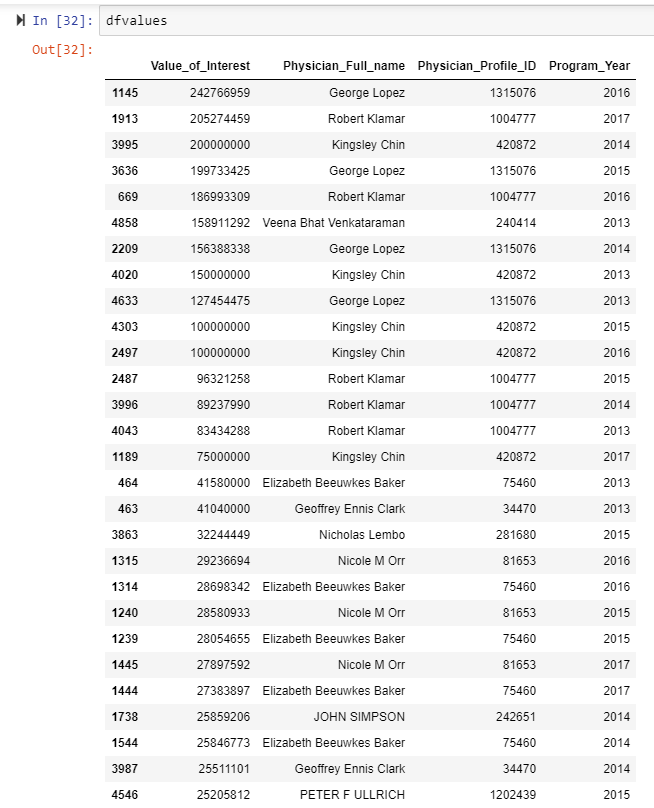


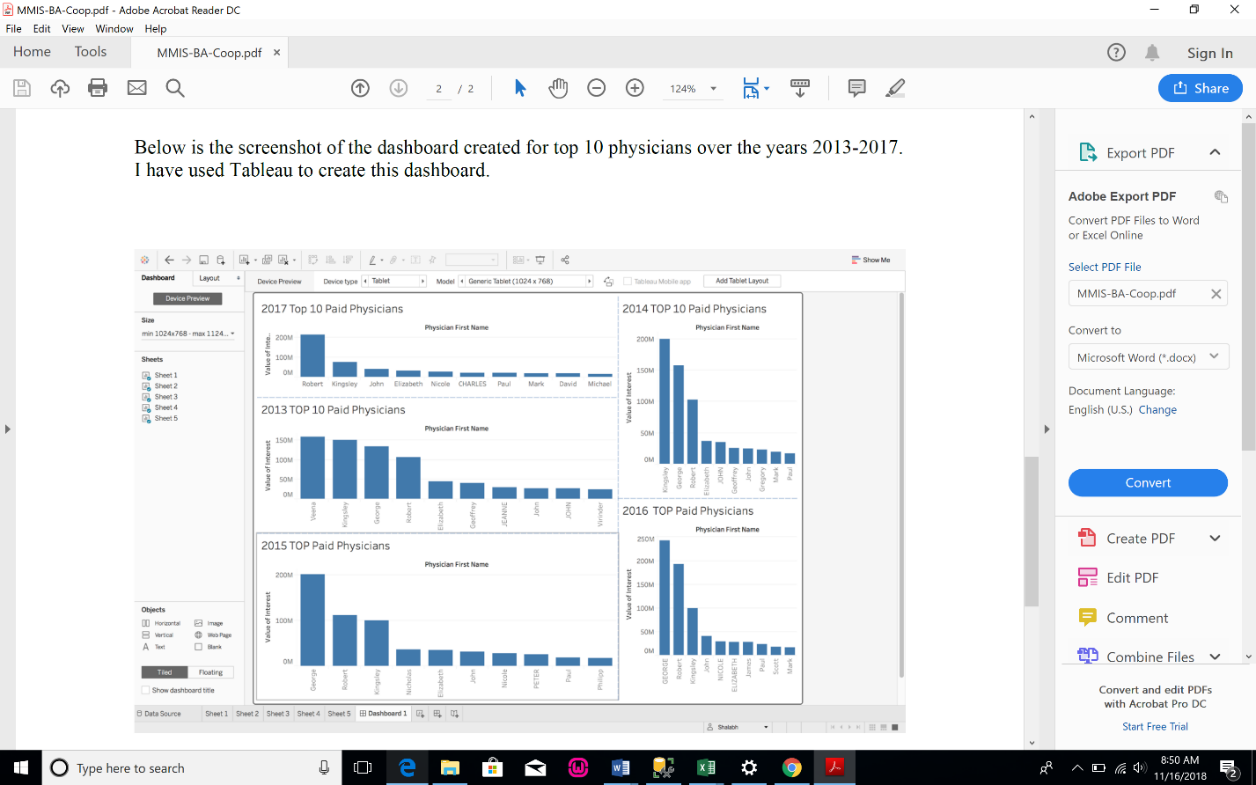
Here we have calculated female beneficiary and female providers for a particular area and in some areas the provider to beneficiary ratio is very less. For the treatment related to female problems we should have reasonable female providers in any particular area for proper treatment. Form the data we can see that there are some areas which have 0 female providers.



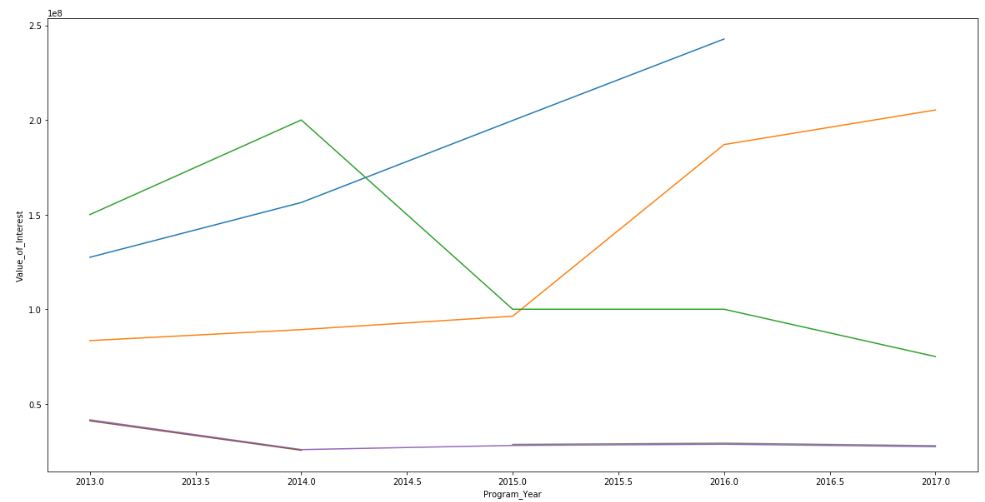
**Open payments Summary Data Set**

We have created a summary table for Open payments. We have determined a list of top physicians who have been paid the highest across the years of 2013 – 2017**.**This will help us to provide the Physicians on need of basis like severity of the case we send or appoint the top most know doctor for the case and for lower severity we can assign the needful doctor.





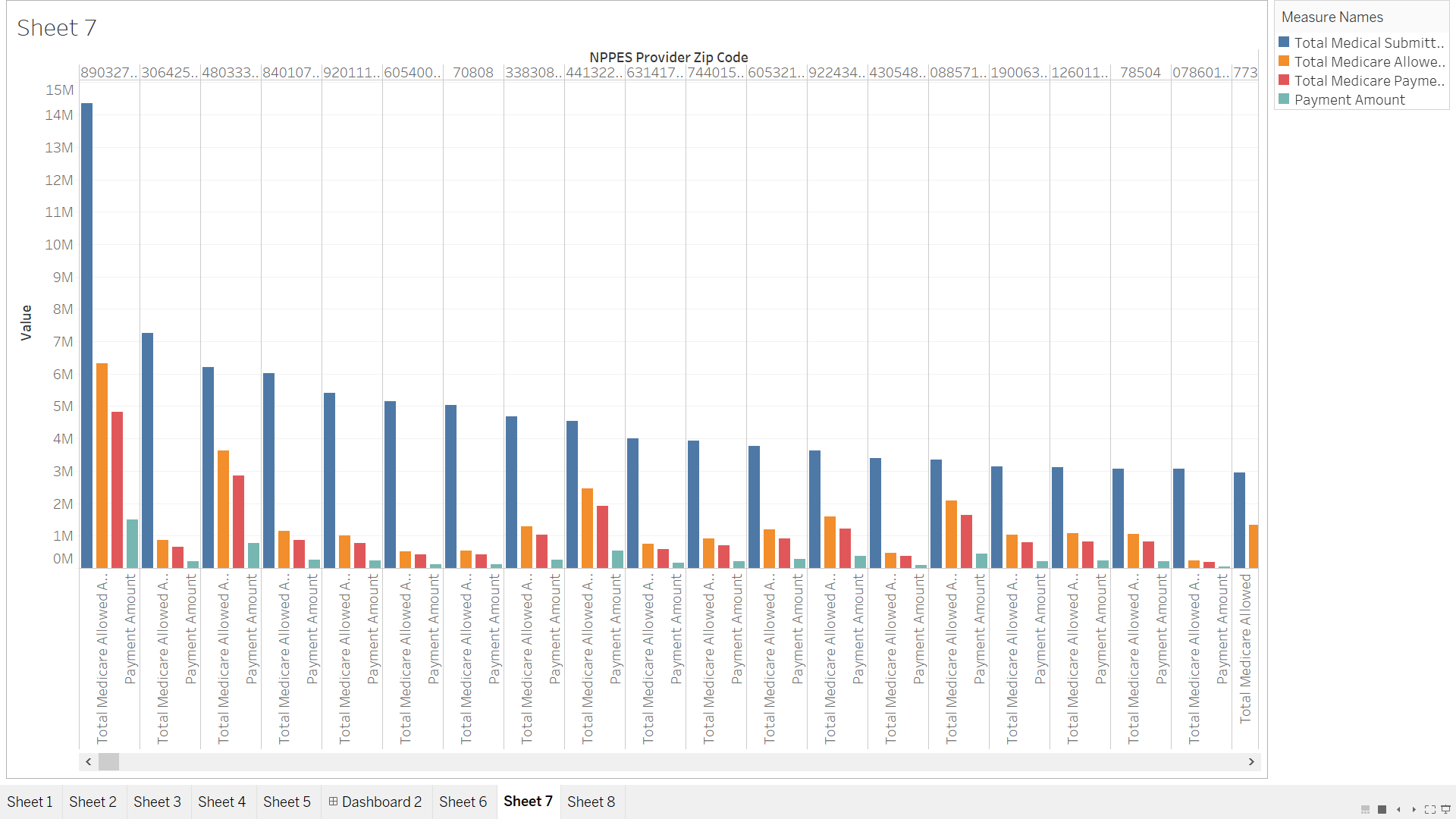
Below is the trend analysis based on the program years from 2013 to 2017, split by 6 months.

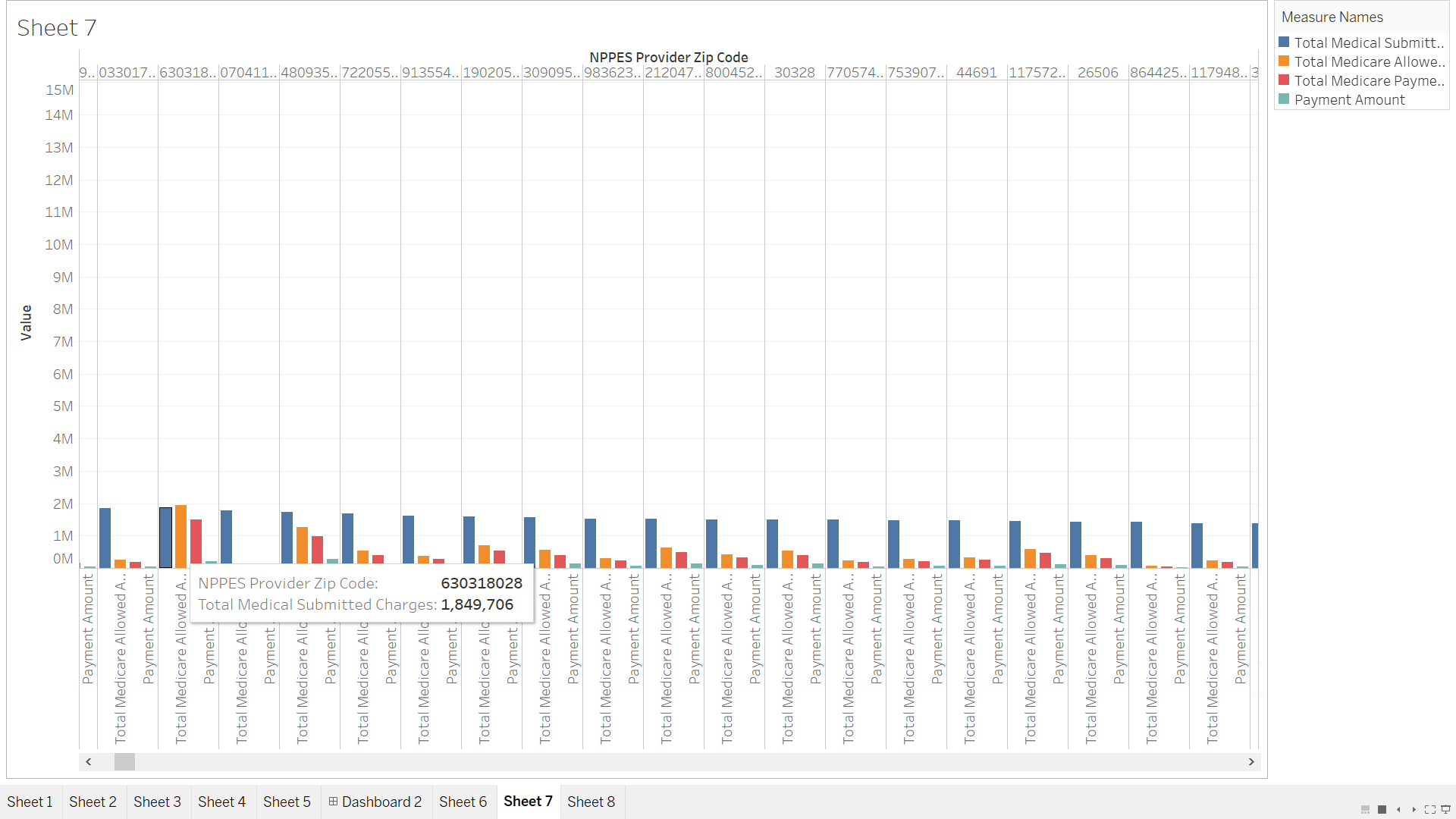


**Visualization Plan**

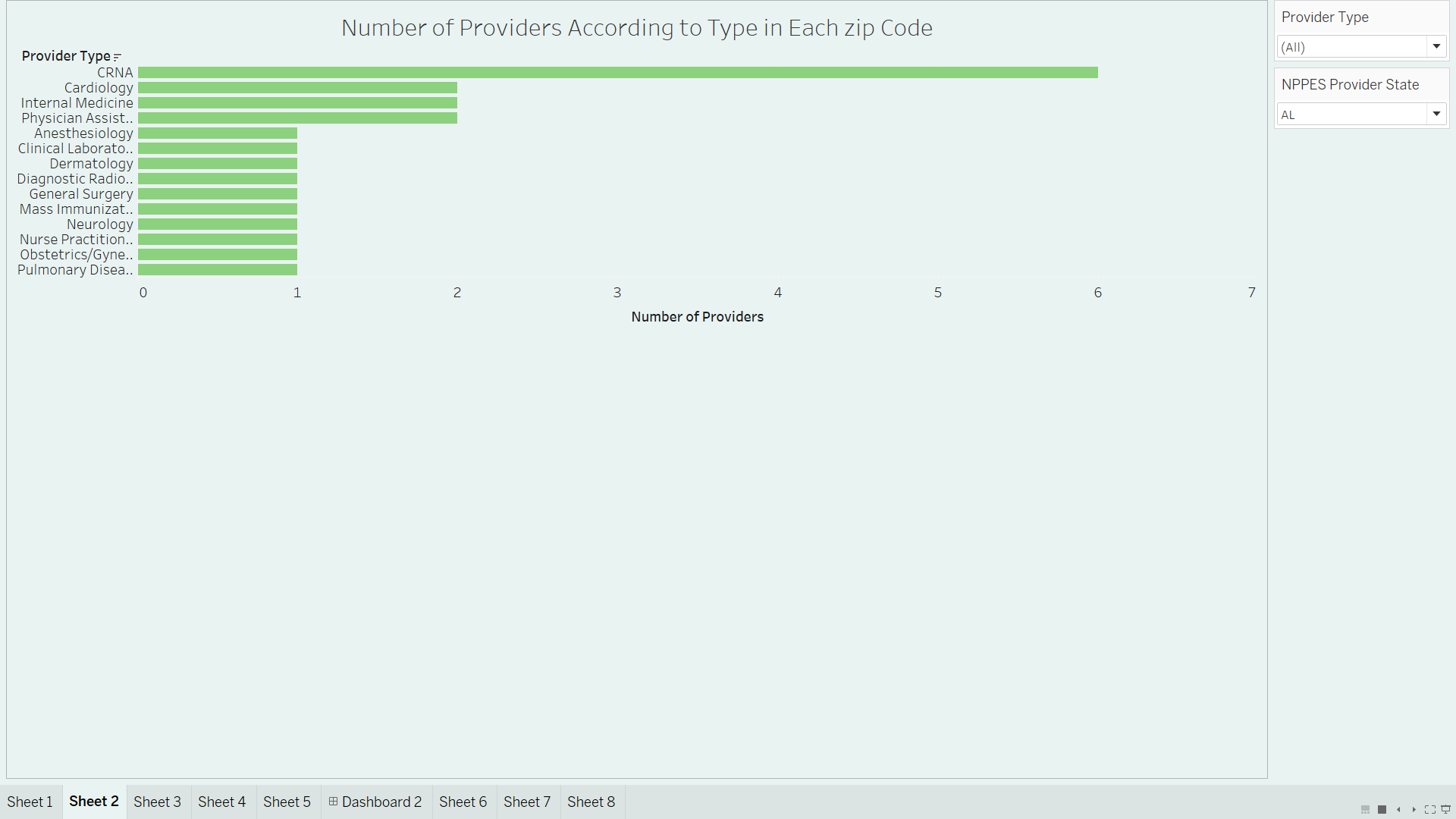
1. We can calculate the percentage of a particular disease in a specific area and the cost of treatment for that disease after that we can relate this data with low income group areas where there is problem of HIPSA which will further help to adjust the total Medicare payment allowed.
2. As of now we have used different tools for visualization like Tableau, Excel, Python. We are planning to use R and SSRS/SSAS as well for further visualization as the data set are very big.
3. We will add more Visualization reports in next draft based on our analysis as we have added most of the reports in this draft.

**Analysis Of Medicare Payments:**

****

****

**Analysis of Providers allocation:**

****