Arshya Srinivas

IEMS 308

January 26th 2020

**Output for SQL code**

1. I first investigated how many countries in the Medicare table are in the US by using the following code. I realized that the US makes up the majority of the countries in the list.

*SELECT COUNT(\*) FROM MedicareTable WHERE nppes\_provider\_country == "US"*

|  |
| --- |
| 9846847 |

2. I was interested to see if there was a difference in the number of credentials in the dataset, however I realized that this difference is not significant enough to look deeper into

*SELECT COUNT(\*) FROM MedicareTable WHERE nppes\_credentials == "I"*

|  |
| --- |
| 14 |

3. I was interested in looking at the gender split of providers in the dataset. It was interesting to notice that the majority of the providers were male; this could be an area of further investigation

*SELECT COUNT(\*) FROM MedicareTable WHERE nppes\_provider\_gender == "F"*

|  |
| --- |
| 2881018 |

*SELECT COUNT(\*) FROM MedicareTable WHERE nppes\_provider\_gender == "M"*

|  |
| --- |
| 6535107 |

4. I looked at the number of Midwestern states in the United states and saw that Midwestern states made up about a fifth of the data set.

*SELECT COUNT(\*) FROM MedicareTable WHERE nppes\_provider\_country == "US" AND (nppes\_provider\_state == "ND" OR nppes\_provider\_state == "SD" OR nppes\_provider\_state == "OH" OR nppes\_provider\_state == "MI" OR nppes\_provider\_state == "IN" OR nppes\_provider\_state == "IL" OR nppes\_provider\_state == "WI" OR nppes\_provider\_state == "MN" OR nppes\_provider\_state == "IA" OR nppes\_provider\_state == "MO" OR nppes\_provider\_state == "KS" OR nppes\_provider\_state == "NE")*

|  |
| --- |
| 2217435 |

5. I also looked deeper into the gender split of providers in Midwestern states, as seen in the code below.

*SELECT COUNT(\*) FROM MedicareTable WHERE nppes\_provider\_country == "US" AND (nppes\_provider\_state == "ND" OR nppes\_provider\_state == "SD" OR nppes\_provider\_state == "OH" OR nppes\_provider\_state == "MI" OR nppes\_provider\_state == "IN" OR nppes\_provider\_state == "IL" OR nppes\_provider\_state == "WI" OR nppes\_provider\_state == "MN" OR nppes\_provider\_state == "IA" OR nppes\_provider\_state == "MO" OR* *nppes\_provider\_state == "KS" OR nppes\_provider\_state == "NE") AND nppes\_gender == “F”*

|  |
| --- |
| 672812 |

*SELECT COUNT(\*) FROM MedicareTable WHERE nppes\_provider\_country == "US" AND (nppes\_provider\_state == "ND" OR nppes\_provider\_state == "SD" OR nppes\_provider\_state == "OH" OR nppes\_provider\_state == "MI" OR nppes\_provider\_state == "IN" OR nppes\_provider\_state == "IL" OR nppes\_provider\_state == "WI" OR nppes\_provider\_state == "MN" OR nppes\_provider\_state == "IA" OR nppes\_provider\_state == "MO" OR nppes\_provider\_state == "KS" OR nppes\_provider\_state == "NE") AND nppes\_provider\_gender == "M"*

|  |
| --- |
| 1454987 |

6. I decided to subset the dataset by removing providers that were not in Midwestern states. I also only included providers that provided the top ten used services.

*DELETE FROM MedicareTable WHERE nppes\_provider\_country != "US"*

*DELETE FROM MedicareTable WHERE nppes\_provider\_state NOT IN (SELECT nppes\_provider\_state FROM MedicareTable WHERE (nppes\_provider\_state == "ND" OR nppes\_provider\_state == "SD" OR nppes\_provider\_state == "OH" OR nppes\_provider\_state == "MI" OR nppes\_provider\_state == "IN" OR nppes\_provider\_state == "IL" OR nppes\_provider\_state == "WI" OR nppes\_provider\_state == "MN" OR nppes\_provider\_state == "IA" OR nppes\_provider\_state == "MO" OR nppes\_provider\_state == "KS" OR nppes\_provider\_state == "NE"))*

*DELETE FROM MedicareTable WHERE provider\_type NOT IN (SELECT provider\_type, COUNT(\*) FROM MedicareTable GROUP BY provider\_type ORDER BY COUNT(\*) DESC LIMIT 10)*