HS 256-F1 Healthcare Data Analytics and Data Mining
Assignment 1 : Group 2

Chitra Pai Kulyadi, Qingqiu Zhang, Ruifeng Liu, Wilfred Gwaikolo, Zhizun Li

Introduction

This paper presents a brief analysis of the National Plan and Provider Enumeration System (NPPES) database which contains information on healthcare providers in the US. The dataset helps provide an insight into the supply side of the US healthcare market. This database is obtained from the Centers for Medicare and Medicaid Services CMS which is a federal agency that oversees major healthcare programs in the US. The CMS provides health coverage to more than 100 million people through Medicare, Medicaid, the Children's Health Insurance Program, and the Health Insurance Marketplace. It seeks to strengthen and modernize the US Nation's health care system, to enable it to provide access to high quality care and improved health.

Methods

Files were downloaded from the CMS data warehouse and inputted into Python for coding and analysis. We used "pandas" to import data and retrieve useful information and used "scipy" to conduct statistical tests. Our analysis focused on specific geographic areas assigned to us which includes 9 states namely Arizona, Georgia, Kentucky, Nevada, Rhode Island, Texas, Vermont, West Virginia, Wisconsin.

Results

Question 1

This section shows the list of doctors for each group member and the state in which said doctor was first licensed. Group members are ordered alphabetically by last name and we used the provider's NPI to show the state in which they were licensed.

a. The healthcare provider of Gwaikolo, Wilfred was first licensed in MA.
b. The healthcare provider of Li, Zhizun was first licensed in MA.
c. The healthcare provider of Liu, Ruifeng was first licensed in MA.
d. The healthcare provider of Pai Kulyadi, Chitra was first licensed in MI.
e. The healthcare provider of Zhang, Qingqiu was first licensed in MA.

Question 2

In this section, we ran a statistical test to explore gender differences in practicing as a "*Sole Proprietor*". Data presented here includes only states that were assigned to group two (Arizona, Georgia, Kentucky, Nevada, Rhode Island, Texas, Vermont, West Virginia, and Wisconsin). We considered only values with Male/Female (M/F) from the gender variable and values with Y/N for Sole Proprietor to run a Fisher's Exact Test for a 2x2 table to test for the gender preference. We excluded the null values during the data cleaning process. Numbers of male and female being a sole proprietor or not are calculated before conducting Fisher's Exact Test. After the test, the result showed a p-value of 0.6369 which is greater than 0.05. This is not statistically significant, therefore we do not reject the Null Hypothesis. In other words, we don't have sufficient evidence to say that there is a significant association between gender and preference for sole proprietorship. The data showed that 32% female, as well as 32% of male had sole proprietorship (Figure 1).

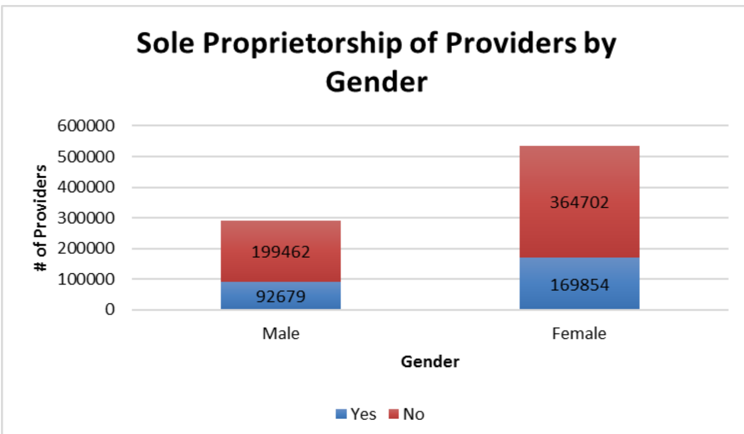

Figure 1. Relationship between sole proprietorship and gender among healthcare providers.

Question 3

In the section, we conducted a statistical analysis to test the hypothesis that male doctors are more likely than their female peers to choose the practices that are associated with higher risk for a higher reward. We did a 2x2 cross-tab (2 practice categories X 2 genders) followed by a Fisher's Exact Test to generate a p value to derive a conclusion.

For this analysis, only data from states assigned to group 2 was considered.We cross-tabulated providers designated as male or female by low and high risk-reward categories. Low risk/reward category included "Obstetrics & Gynecology" and "Pediatrics" and the high risk/reward category included "Surgery" and "Orthopaedic Surgery". We found the *Health Care Provider Taxonomy Code Set* for all four categories.

	# High risk practices performed	# Low risk practices performed
By male	8097	7151
By female	1651	11611

Table 1. Providers with high risk and low risk practices performed based on gender.

As shown in the table above, among the 4 mentioned practice categories, male doctors performed 53.1% of high risk practices, whereas the ratio for female doctors is only 12.4%. To test if there is a statistically significant difference between male and female doctors in choosing practices with different risk levels, we conducted the Fisher Exact test for the following Hypothesis:

H_0: *Males are less or equally likely to do high risk practices, compared to their female peers.*

H_a: *Males are more likely to do high risk practices, compared to their female peers.*

Based on the analysis, we concluded that in the designated states of Arizona, Georgia, Kentucky, Nevada, Rhode Island, Texas, Vermont, West Virginia and Wisconsin, there was an observed imbalance that is statistically significant ($p = 0.00 < 0.001$); males prefer to choose the high-risk practices while females prefer to do low-risk practice.

Question 4

We aim to find the density of Magnetic Resonance Imaging (MRI) centers per state. Thus, after filtering it to obtain healthcare facilities in the national database, we found the number of MRI centers setting the code "*Healthcare Provider Taxonomy Code_1*" variable equal to 261QM1200X. We found MRI centres using the code "*Healthcare Provider Taxonomy Code_1*" variable in the national database after filtering it to obtain healthcare facilities.

In order to calculate the MRI density (per 1,000,000 population) for all of the states, we looked up data for US population by states on the United States Census Bureau. We used US Census Bureau population estimates as of July 1, 2021 for the population statistics.(https://data.census.gov/cedsci/table?tid=PEPPOP2021.NST_EST2021_POP)

We calculated the MRI density (per 1,000,000 population) for all of the states. We found the MRI density to be highest in the states of Texas, Wyoming and Florida, with Florida having the highest density of MRIs per 1,000,000 population compared to all other states. Table 1 shows the ranked MRI density across states. We visualized MRI density across states in the US by graphing a heatmap on Tableau. (Figure.1)

Rank	State	MRI Density
1	FL	13.93
2	WY	13.87
3	TX	8.71
4	PR	7.91
5	IL	7.10
6	NJ	7.00
7	NH	6.53
8	ND	6.42
9	MA	6.26
10	GU	5.92

Rank	State	MRI Density
27	MD	3.24
28	IA	3.13
29	CO	3.12
30	KS	3.06
31	CA	3.06
32	WI	3.05
33	DC	2.90
34	GA	2.89
35	KY	2.89
36	AK	2.73

11	ME	5.87
12	OK	5.81
13	LA	5.80
14	MT	5.53
15	AR	5.31
16	MN	5.26
17	DE	5.05
18	MI	4.66
19	AL	4.18
20	ID	3.81
21	NM	3.78
22	MO	3.74

37	MS	2.70
38	CT	2.50
39	OR	2.36
40	NY	2.33
41	PA	2.15
42	TN	2.03
43	SC	1.95
44	NV	1.93
45	VA	1.85
46	WA	1.56
47	VT	1.56
48	UT	1.53

23	OH	3.64	49	NC	1.44
24	NE	3.57	50	SD	1.13
25	IN	3.54	51	HI	0.69
26	AZ	3.50	52	WV	0.56

Table 2. Density of MRIs in different states in the US

The high MRI density noted in Florida could be due to the fact that it is one among the ten most populated states in the US . Also, Florida has the second highest number of Americans of age 65 or older. The presence of a larger number of aged individuals in the state could predispose them to increasing investment in medical equipment that can be utilised to enhance diagnostic and treatment modalities in the clinical setting.(Killduff, 2021)

In addition, Florida has the second highest number of medical device manufacturing companies in the country (https://business.orlando.org/l/medical-technology/) which could explain why the density of MRI centers and facilities is high. Fast growing medical device companies have set up a market in Orlando, Florida because of its recognized medical device training expertise and expansive workforce, making the area one of the top cities for medical device companies.

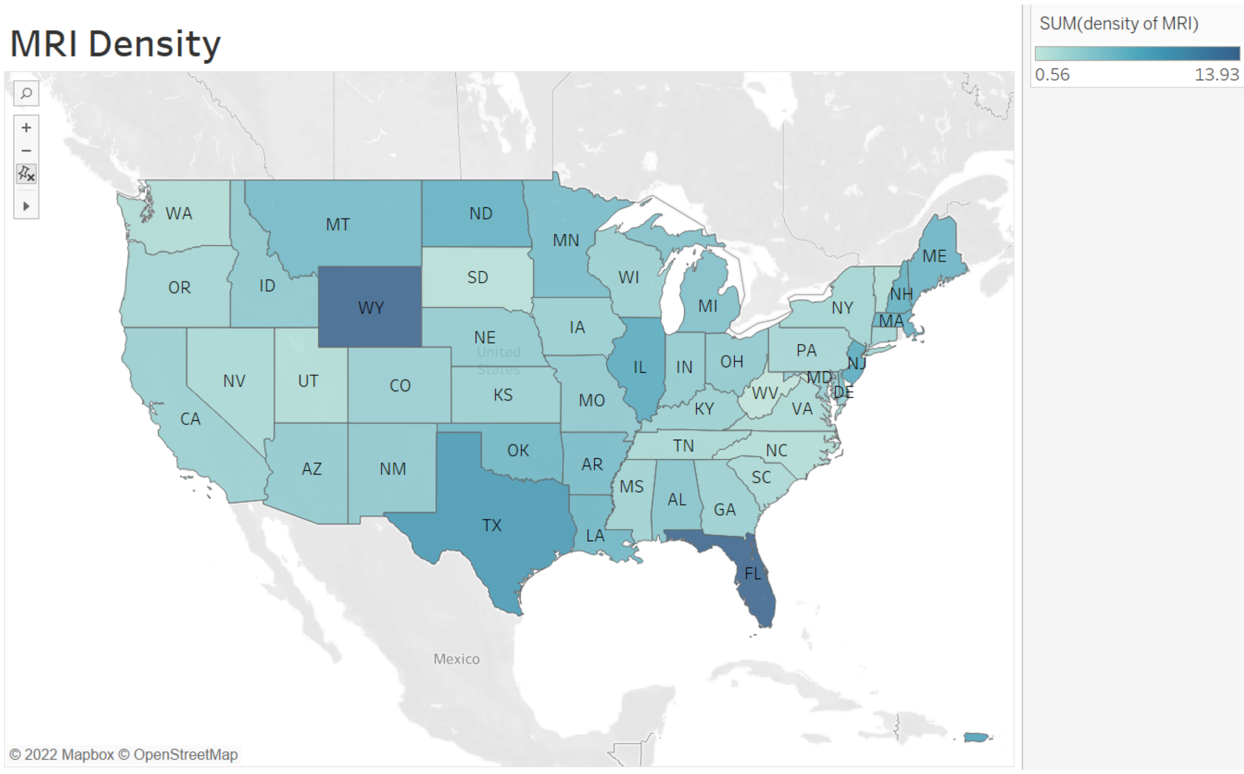

Figure 1: graphic visualisation of MRI density per 1,000,000 population in different states. The darker states have higher MRI densities.

References

Killduff, L. (2021, December 22). *Which U.S. states have the oldest populations?* PRB. Retrieved January 30, 2022, from https://www.prb.org/resources/which-us-states-are-the-oldest/

Prince, C. J., Keller, M., Braun, J., & Roche, A. (2021, September 30). *Medical Technology: Orlando Economic Development*. Orlando Economic partnership. Retrieved January 30, 2022, from https://business.orlando.org/l/medical-technology/

United States Census Bureau. (n.d.). NST_EST2021_POP Annual Estimates of the Resident Population for the United States, Regions, States, District of Columbia, and Puerto Rico: April 1, 2020 to July 1, 2021. Retrieved January 30, 2022, from https://data.census.gov/cedsci/table?tid=PEPPOP2021.NST_EST2021_POP