Clustering: How does Medicare Payment vs. the Amount Charged for Service Vary by State?

Anna Blakley
IEMS 308
Professor Klabjan
January 27, 2020

Executive Summary

When examining *Medicare Physician and Other Supplier Data CY 2017* provided by the Center for Medicare and Medicaid Services, the greatest dependency and variation identified were between the Average Submitted Charge Amount (average_submitted_chrg_amt) and the Average Medicare Payment Amount (average_Medicare_payment_amt). This dependency lead to the business question that guided my clustering moving forward: **How does Medicare Payment vs. the Amount Charged for Service Vary by State?** To answer this question clustering was executed on four separate subsections of the original data using Python.

The first subsection (1) that was clustered was the entirety of the dataset grouped by state. However, further exploration of the data revealed that both variables depended on other variables within the data set that lead to skewed values in the clustered data. The next subset (2) that was clustered included only female, M.D.s who practice Diagnostic Radiology who are working in a facility that participates in Medicare. The data was then grouped by individual (NPI) and state. This data was again skewed. This new subset was the same as the previous but filtered to include only HCPCS code "71010." The data was clustered at this point (3) and clustered again (4) once it was grouped by state. The third clustered subset revealed very little about the data because there were no "clean" clusters as many states appeared in multiple clusters. For this reason, the final subsection (4) was used to answer the business question.

The fourth subsection was clustered into six different clusters. Cluster four only has one state, but will remain in the data because it does not fit within the other clusters and the submitted, payment, and percentage means do not flag it as an outlier. These clusters are categorized as follows:

- **Low average submitted charge amount, but high payment by percentage** (Cluster 0) - States: North Dakota, Missouri, Michigan, Utah, Louisiana, South Dakota
- **Low average submitted charge amount, but moderate payment by percentage** (Cluster 5) - States: Oklahoma, Indiana, Kentucky, Nevada, New Mexico, Minnesota, Tennessee, West Virginia, Colorado, Nebraska, Arizona, Arkansas, Alabama, Kansas, Idaho
- **Moderate average submitted charge amount and moderate payment by percentage** (Cluster 3) - States: South Carolina, Oregon, Ohio, Vermont, Texas, Mississippi, Wyoming, Montana, California, Georgia, Hawaii, Iowa, North Carolina, Illinois
- **Low average submitted charge amount, but high payment by percentage** (Cluster 1) - States: Maryland, Massachusetts, New York, Pennsylvania, Rhode Island, Florida, Delaware, Connecticut, Virginia.
- **High average submitted charge amount, but moderate payment by percentage**
- (Cluster 4) - States: Alaska
- **High average submitted charge amount, but low payment by percentage** (Cluster 2) - States: New Hampshire, Maine, Washington, New Jersey, Wisconsin.

States falling in cluster 0 receive the greatest percentage of their charges from Medicare, but also submit the lowest charges. The amount paid by Medicare can be reduced in these states to match the others. The states in cluster 2 submit the highest charges and proportionally receive the least amount from Medicare. These states can continue receiving the same payment from Medicare because it is equivalent to that received by other states, but should lower charges to match other states.

Summary

Business Question: How does Medicare Payment vs. the Amount Charged for Service Vary by State?

Process:

1. Understand and explore the dataset.
2. Identify a business question.
3. Remove outliers.
4. Execute clustering on entirety of the dataset grouped by state using the mean of the submitted charge, the Medicare payment received, and the percent of the request that was paid by Medicare (payment/ submitted charge).
5. Filter dataset to include only female M.D.s working in a facility that participates in Medicare and practices Diagnostic Radiology, then grouped by NPI and state and cluster
6. Filter the same subsect used in 4 to include only HCPCS code "71010" and cluster
7. Group the subsect from 5 by state.
8. Examine the clusters created in 6.
9. Identify feature of the clusters.

Exploration

Dataset features:

26 variables and 9,847,444 rows of data
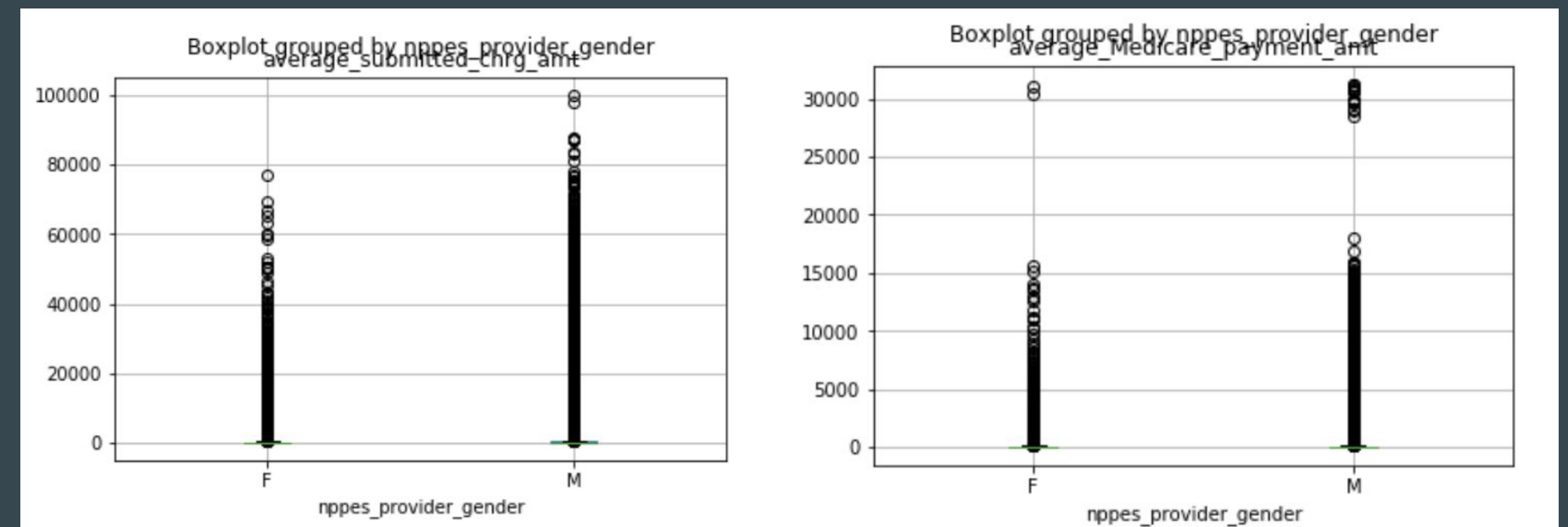
Figures 1 and 2: Exploring effect of provider gender on the submitted charge and Medicare payment
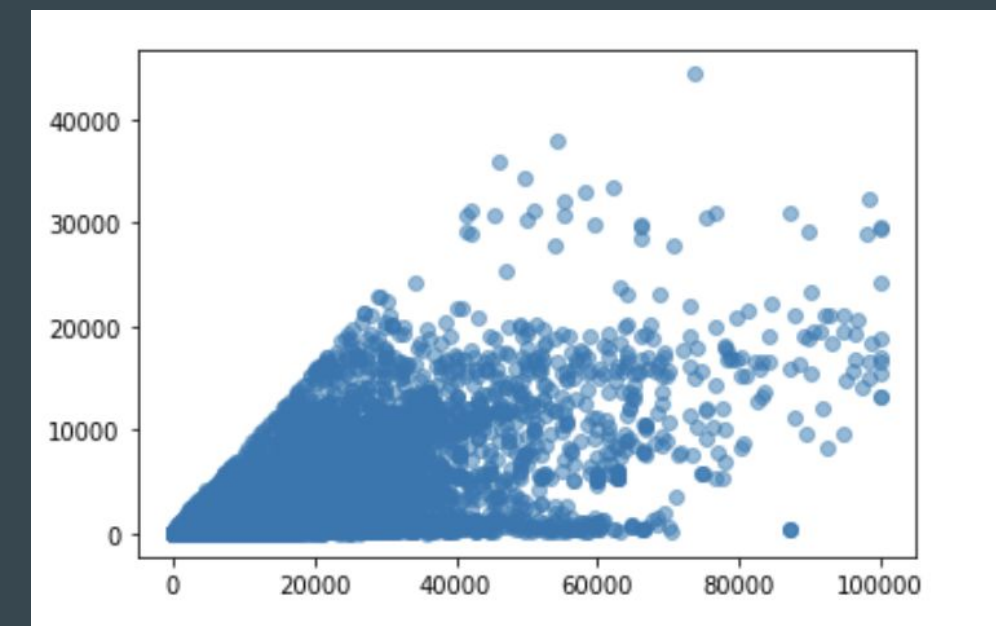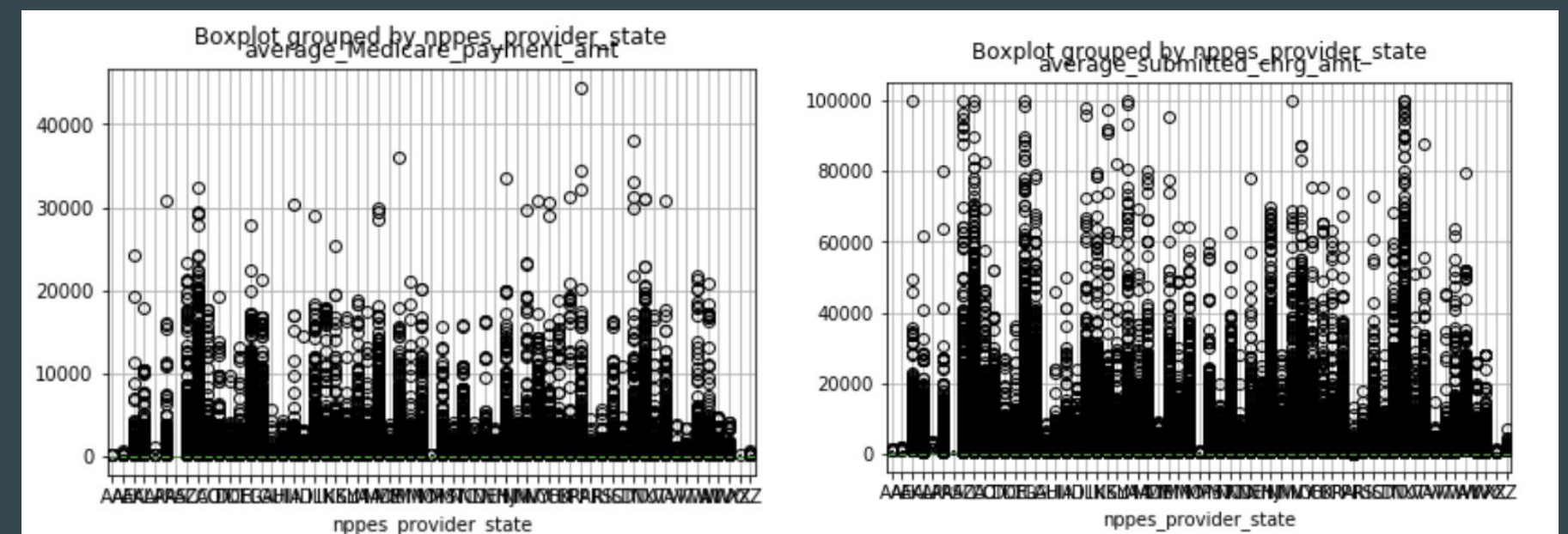

Figure 3: Exploring the relationship between submitted charge and Medicare payment

Figures 4 and 5: Exploring effect of provider state on the submitted charge and Medicare payment

Initial Clusters and Learnings

1st Clustering: Grouping the entire data set by state and clustering the the mean of the submitted charge, the payment received, and the percent paid. (see cluster breakdown in appendix)

Key Learning: Other variables affect the submitted charge, so some states have higher values because more expensive procedures are done in those states

2nd Clustering: Including only female, M.D.s who practice Diagnostic Radiology, work in a facility that participates in Medicare then grouping by NPI and then by state. The same data was used in clustering(see cluster breakdown in appendix)

Key Learning: Procedure type is the main variable that inflates the values, so this variable must be filtered.

3rd Clustering: Including only female, M.D.s who practice Diagnostic Radiology, work in a facility that participates in Medicare and only HCPCS code 71010. The same data was used in clustering(see a sample of the cluster breakdown in appendix)

Key Learning: There were no "clean" clusters as many states appeared in multiple clusters. This is likely because providers within a state charge different amounts for the same procedure. To answer the business question I will group by state as I did in the previous clusters.

Final Clustering

Final Cluster: Including only female, M.D.s who practice Diagnostic Radiology, work in a facility that participates in Medicare and only HCPCS code 71010 and then grouped by state. The same data was used in clustering, where k = 6.
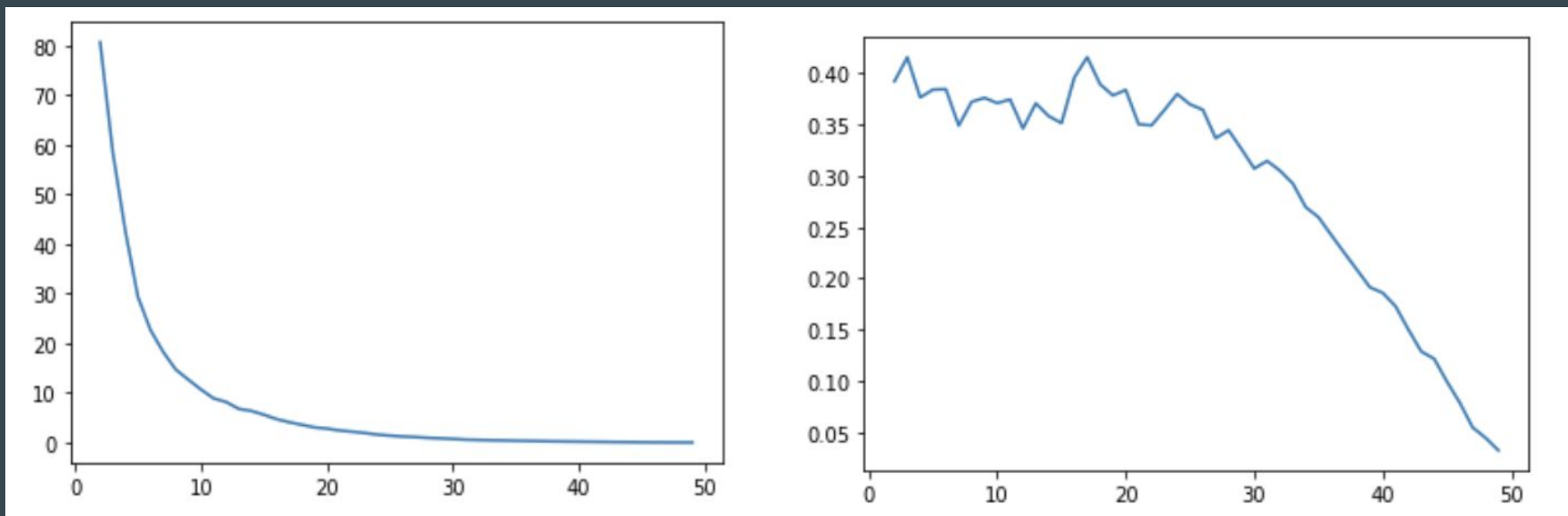
Figures 6 and 7: Plot of k value vs inertia and silhouette plot

Cluster	Request	Payment	Percentage	Count
0	29.332817	6.990433	0.238609	6
1	37.131949	7.334442	0.198087	9
2	51.261256	7.116773	0.139352	5
3	40.45463	6.905439	0.171414	14
4	48.333188	8.68436	0.179677	1
5	32.060014	6.7	0.209482	15

Figure 8: Mean values and counts for each of the final clusters

Cluster 4 only included one state, however, this cluster does not fit cleanly into the other clusters when comparing the mean values. Additionally, the values for cluster four are not vastly different than those of the other clusters.

Business Insights

To best understand which states submit large bills and which states receive high Medicare payments I created three categories to interpret the findings: low, moderate and high.

Cluster	Request	Category	Percent	Category
0	29.332817	Low (<35)	23.8609	High (21+)
1	37.131949	Moderate (35-45)	19.8087	Moderate (16-21)
2	51.261256	High(45+)	13.9352	Low (<16)
3	40.45463	Moderate (35-45)	17.1414	Moderate (16-21)
4	48.333188	High (45+)	17.9677	Moderate (16-21)
5	32.060014	Low (<35)	20.9482	Moderate (16-21)

Figure 9: Categorization of clusters
This categorization can be interpreted as:

- Low average submitted charge amount, but high payment by percentage (Cluster 0)
- Low average submitted charge amount, but moderate payment by percentage (Cluster 5) Moderate average submitted charge amount and moderate payment by percentage (Cluster 3)
- Low average submitted charge amount, but high payment by percentage (Cluster 1)
- High average submitted charge amount, but moderate payment by percentage (Cluster 4) High average submitted charge amount, but low payment by percentage (Cluster 2)

Meaning: States falling in cluster 0 receive the greatest percentage of their charges from Medicare, but also submit the lowest charges. The amount paid by Medicare can be reduced in these states to match the others. The states in cluster 2 submit the highest charges and proportionally receive the least amount from Medicare. These states can continue receiving the same payment from Medicare because it is equivalent to that received by other states, but should lower charges to match other states.

Value: Clustering using other HCPCS codes should be executed to verify these clusters, but overall these results can be used to identify which states Medicare can provide a lower payment and which states should lower prices of services.

Appendix

Clustering 1: Grouped by States

state	request	payment	percentage	Cluster	state	request	payment	percentage	Cluster
AL	271.6082117	66.99192818	0.246649127	0	AZ	369.2623142	88.94133667	0.240862209	1
AR	262.5852527	67.52815127	0.25716658	0	CA	399.8497418	91.55744163	0.228979619	1
HI	275.6103768	71.15696546	0.258179559	0	CO	364.4746189	81.07344337	0.222439202	1
IA	276.5085439	63.8277325	0.230834576	0	DE	380.5665038	86.7703841	0.228003209	1
ID	273.9939675	66.90862294	0.244197431	0	FL	394.6506143	89.96532065	0.227961943	1
KS	312.1704403	73.79918587	0.236406707	0	MD	355.4493604	91.63646222	0.257804549	1
KY	267.658589	65.99840972	0.246576842	0	NJ	409.3505805	85.62635391	0.20917609	1
ME	258.8773022	65.3213614	0.252325564	0	NY	373.8709603	80.61053726	0.215610587	1
MI	296.5398809	74.31644407	0.250611971	0	WI	484.6180987	64.22527129	0.132527595	2
MN	269.8093715	64.45426099	0.238888148	0	CT	356.6148552	77.33026608	0.21684533	3
MT	275.0770328	74.38796981	0.270425957	0	GA	385.3306495	76.0750088	0.197427869	3
ND	273.8193733	64.92136255	0.237095578	0	IL	381.0209301	77.68827025	0.20389502	3
NE	291.1356169	67.48877751	0.231812164	0	IN	357.206391	74.33182569	0.208092093	3
NM	314.6512396	70.75422471	0.224865552	0	LA	362.7825307	74.86892164	0.206374109	3
OH	308.3541905	71.87505262	0.233092511	0	MA	325.9588699	73.19652041	0.224557535	3
OK	302.4925023	74.47557843	0.246206362	0	MO	343.6529379	75.4698232	0.219610586	3
OR	312.4959531	74.6894702	0.2390094	0	MS	336.4871123	66.79184598	0.198497486	3
PA	314.3182031	76.25617031	0.24260819	0	NC	323.4745363	68.30624931	0.211164224	3
RI	279.1622186	69.99845307	0.250744723	0	NH	408.2480308	72.61900391	0.177879618	3
SD	289.8530555	68.62183242	0.236746969	0	SC	345.0610917	71.67711404	0.207722968	3
UT	297.7199	77.29292372	0.259616249	0	TN	313.3957465	68.21465516	0.21766299	3
VA	323.5386117	75.2421876	0.232560149	0	TX	413.8534004	79.70658656	0.192596186	3
VT	280.0301686	61.79433388	0.220670273	0	WY	328.5652845	64.70540202	0.196933167	3
WA	296.4660425	75.33491715	0.254109768	0	AK	601.6852912	93.66055551	0.155663695	4
WV	279.4961774	64.83790232	0.231981356	0	NV	480.2014423	92.74309393	0.193133726	4

Cluster	request	payment	percentage
0	288.158889	69.930969	0.242935
1	380.934337	87.02266	0.228855
2	484.618099	64.225271	0.132528
3	355.832312	72.92725	0.205661
4	540.943367	93.201825	0.174399

Upper: Clustering results
Lower: Mean data for each cluster

Clustering 2: Filtered and Grouped by States and NPI

state	request	payment	percentage	Cluster	state	request	payment	percentage	Cluster
AR	130.6182776	29.15850916	0.223234525	0	CT	158.2576797	32.49053757	0.205301491	3
DE	105.47953	21.58923749	0.204677035	0	FL	179.3418873	34.17347848	0.190549341	3
ID	100.887345	29.18130835	0.289246469	0	IN	159.9319609	31.69651814	0.198187516	3
KS	131.458248	28.47657111	0.21662065	0	MI	141.5174883	31.01814948	0.219182448	3
ND	104.3333333	27.91489225	0.267554878	0	MN	135.8581465	32.69672566	0.240668127	3
SD	123.9305224	30.23025538	0.243929056	0	MO	159.0434033	30.85393872	0.193996972	3
WV	127.3697326	27.54329456	0.21624678	0	NE	158.4785485	30.69271762	0.193671118	3
CA	237.9760072	33.13208232	0.139224465	1	NJ	169.9635256	32.77227993	0.192819488	3
NH	267.6992894	30.83793421	0.115196175	1	NV	152.1079977	32.90440895	0.216322675	3
OH	230.1393495	33.63062126	0.146131556	1	OK	154.6726682	30.14868761	0.194919296	3
TX	242.2670752	34.01514467	0.140403497	1	PA	172.7825408	32.9869763	0.190916143	3
VT	250.9319058	34.29927923	0.136687597	1	TN	164.374536	30.76563206	0.187167872	3
WI	343.6524895	35.28277412	0.102669921	1	UT	143.2051694	32.93022477	0.229951369	3
HI	186.0926955	31.05523539	0.166880464	2	VA	179.39467	34.12664227	0.190232197	3
IA	146.9897283	26.86901297	0.182795174	2	WA	143.7038331	30.31379077	0.210946292	3
LA	165.4115408	30.18998285	0.182514368	2	AK	273.8040815	42.15250858	0.153951352	4
ME	161.6200981	25.71299988	0.159095312	2	AL	203.4900527	38.83424346	0.190840992	4
MT	201.0853967	29.58587734	0.14713091	2	GA	208.0270999	35.49957953	0.170648822	4
NC	177.5562554	29.65285917	0.167005432	2	IL	222.2836321	33.99172692	0.152920512	4
NM	167.376754	29.15438731	0.174184208	2	KY	207.8386721	36.91560077	0.177616612	4
OR	147.1321187	27.05686439	0.183895023	2	MA	191.0554482	35.72015514	0.186962243	4
RI	145.7137884	28.03250667	0.192380604	2	MD	199.0080279	34.52728712	0.173496956	4
WY	182.6705142	24.71146045	0.135278868	2	MS	222.7059496	35.80906048	0.160790767	4
AZ	141.7227341	30.38413873	0.214391424	3	NY	228.2284761	35.260487	0.15449644	4
CO	144.1107384	30.98675247	0.215020427	3	SC	203.7884627	33.69225158	0.165329534	4

Cluster	request	payment	percentage
0	117.725284	27.727724	0.237358
1	262.111019	33.532973	0.130052
2	168.164889	28.202119	0.169116
3	156.380443	31.878918	0.204956
4	216.02299	36.24029	0.168705

Upper: Clustering results
Lower: Mean data for each cluster

Clustering 3: Filtered and One HCPCS Code

request	payment	percent	cluster	state	request	payment	percent	cluster	state
35	7.69013913	0.219718261	0	AZ	23	7.079787234	0.307816836	6	PA
34.90677966	7.905762712	0.226482156	0	CA	20	6.783542857	0.339177143	6	PA
26.74657534	7.32630137	0.273915493	0	CA	22	6.648943089	0.302224686	6	TN
30	7.585528289	0.252850943	0	CA	23	7.405	0.321956522	6	TX
23	6.645726291	0.288944621	1	AL	19.12012012	7.16033033	0.374491911	6	UT
33	6.882089286	0.20854816	1	AL	76	6.608	0.086947368	7	WI
25.47118891	6.540423049	0.256777297	1	AR	76.63636364	6.733818182	0.087867141	7	WI
27.95098039	6.817843137	0.243921431	1	AZ	60.33825397	6.882063492	0.114058048	7	WY
23	6.26	0.272173913	1	AZ	28.95	6.153888889	0.212569564	8	AL
29	7.18	0.247586207	1	CA	17	4.563484849	0.268440285	8	AR
45.00693241	6.721923744	0.149353075	2	AL	34	6.098716418	0.179374012	8	AR
38	7.067827715	0.185995466	2	AR	34	6.384940239	0.18779236	8	AR
10.98958333	6.777604167	0.616729858	4	SC	43.82333333	5.792	0.132167034	8	AZ
9.609884793	6.786382489	0.706187705	4	SD	36.03212435	6.200051814	0.172070116	8	AZ
14	7.091834532	0.506559609	4	TX	44.82	8.124150943	0.181261735	9	NY
14	7.11	0.507857143	4	TX	58.53015873	8.247333333	0.140907414	9	OH
33	7.35	0.222727273	5	AR	62.91674462	7.534911132	0.119760029	9	OK
38	7.24	0.190526316	5	AZ	50	8.208	0.16416	9	PA
38.94950732	7.163035952	0.183905688	5	AZ	57.71428571	7.447806122	0.129046146	9	PA
31.30357143	7.237321429	0.231197946	5	CA	58.4	7.594913793	0.130049894	9	PA

Upper: Sample of clustering results

Cluster	request	payment	percentage	AK	AL	AR	AZ	CA
0	30.29802375	7.630210595	0.254947586	0	0	0	0.004219409	0.223628692
1	28.55873076	6.85163125	0.241765603	0	0.012315271	0.012315271	0.012315271	0.083743842
2	41.78996388	6.831832823	0.166315329	0	0.015748031	0.002624672	0.026246719	0.083989501
3	253.2931834	7.487026052	0.02953257	0	0	0	0	0.714285714
4	11.7060499	6.996034248	0.610628224	0	0	0	0.068965517	0.137931034
5	39.58496223	7.382464194	0.18882586	0	0	0.002816901	0.014084507	0.18028169
6	21.62623021	7.049046331	0.330697007	0	0.015873016	0	0.047619048	0.055555556
7	75.70847347	6.987835367	0.094795625	0	0	0	0	0.068027211
8	35.42413285	5.906699676	0.176409432	0	0.010989011	0.032967033	0.032967033	0.175824176
9	50.80731738	7.97624404	0.160647558	0.020979021	0	0	0	0.174825175

Sample mean data for each cluster

Final Clustering: Filtered and One HCPCS Code Grouped by State

state	request	payment	percentage	Cluster	state	request	payment	percentage	Cluster
ND	29	7.379032258	0.2544493882	0	MS	38.57607689	6.643612322	0.1722210462	3
MO	29.89143108	6.907421824	0.2310836776	0	WY	43.11275132	6.841335979	0.1586847457	3
MI	30.94059046	6.955523199	0.2248025359	0	MT	42.48007614	7.006382326	0.1649333749	3
UT	28.32254281	6.859645039	0.2421973579	0	CA	44.10472518	7.220955495	0.163722945	3
LA	30.19555512	7.046651775	0.2333671876	0	GA	41.21965465	6.991734478	0.1696213745	3
SD	27.64678452	6.794325521	0.245754638	0	HI	37.08673469	6.794388617	0.1832026646	3
MD	37.06023272	7.385015241	0.1992706116	1	IA	36.73357548	6.715729602	0.1828226497	3
MA	35.64710729	7.3489985	0.2061597436	1	NC	36.71817995	6.867785865	0.1870404762	3
NY	38.5238345	7.563463242	0.1963320458	1	IL	44.02406797	7.266730062	0.1650626668	3
PA	37.92080911	7.185848974	0.1894961933	1	AK	48.33318765	8.684359824	0.1796769517	4
RI	39.67505105	7.260684188	0.1830037768	1	OK	33.96283092	6.82394185	0.2009238236	5
FL	37.98225784	7.085828999	0.1865562871	1	IN	31.76805559	6.755977411	0.2126657513	5
DE	33.36862685	7.467675047	0.2237932979	1	KY	31.34703267	6.787626481	0.2165317066	5
CT	38.62686387	7.428870344	0.1923239321	1	NV	34.33871697	6.65234308	0.1937271881	5
VA	35.38275759	7.283594288	0.205851516	1	NM	31.64869565	6.861214361	0.2167929584	5
NH	55.90669016	7.025531692	0.1256653126	2	MN	33.36664706	6.811650471	0.2041454888	5
ME	48.3380315	7.034522361	0.1455276962	2	TN	33.74449161	6.614152309	0.1960068738	5
WA	51.11628899	7.075434159	0.1384183848	2	WV	30.82401775	6.759983226	0.2193089583	5
NJ	48.3494476	7.45279105	0.1541442854	2	CO	32.93727292	6.974333529	0.2117459313	5
WI	52.59582361	6.995583629	0.1330064471	2	NE	32.25461759	6.45803482	0.2002204739	5
SC	38.05943115	6.771533561	0.1779199887	3	AZ	33.73998786	6.928649958	0.2053542516	5
OR	37.11185617	6.954141396	0.1873832816	3	AR	29.34946294	6.462636615	0.2201960775	5
OH	42.91088284	6.75798718	0.157488887	3	AL	33.71120946	6.756634187	0.2004269291	5
VT	44.53866339	6.968648946	0.1564629115	3	KS	30.32716408	6.646264237	0.2191521838	5
TX	39.68814494	6.87518083	0.1732300877	3	ID	27.58	6.206556542	0.2250383083	5

Clustering results