# IBM Data Science Capstone Project:

# Healthcare Analysis in Seattle Washington

Jeffrey Guest

June 24, 2020

## I. Introduction:

Healthcare delivery models are undergoing substantial changes given evolving reimbursement methodologies. As part of this process, patients are increasingly responsible for making financial decisions about their own health care as ever larger portions of healthcare costs are now the responsibility of the individual patient. One downstream impact of this change in the healthcare marketplace is that individuals seeking health treatment are behaving more and more like consumers of other types of goods and services, particularly when it comes to drivers such as quality and convenience.

To meet the demands of savvy health consumers, in the past decade there has been a precipitous rise in so-called "doc-in-a-box" clinics and urgent-care centers which operate differently than either the traditional physician office or the acute-care hospital emergency room.

Increasingly, understanding healthcare access within communities is a critical component to planning. Public health officials, along with health plan and hospital administrators, and physician practice managers will want to understand the healthcare resources better as they plan for their own access needs. Especially with recent development of COVID-19, there has been an increased awareness around how interconnected health systems are to each other.

## II. Business Problem:

This project intends to use data from both the Washington State Department of Health government data, along with Foursquare API technology to analyze and understand some key elements about both hospitals and other health services such as physician practices. With relation to the IBM Data Science capstone project, this is intended more as an exercise in the types of analysis that can be enhanced with the use of technological tools such as Python pandas, k means clustering, and graphical representation of key charge data.

Making such data understandable and available to public via various web pages, etc. can help drive consumer behavior to more efficient and appropriate care settings at lower overall cost to society.

## III. Data:

The data was taken from the Washington State Department of Health and includes a breakdown of payer census metrics from 1/1/2019 to 12/31/2019; the census information includes information on facility type (acute care, swing bed, etc), total discharges, total patient days, total charges and includes payername (which is a general category describing funding sources for the discharge (e.g. Medicare, Medicaid, Commercial payor, workers compensation or self-pay).

Additionally, zip code mapping associated with neighborhoods was scraped from multiple municipal websites for the city of Seattle.   This data was used to in conjunction with Foursquare API data to complete a k-means analysis grouping of healthcare services within the city of Seattle.  The analysis was confined to the city limits of Seattle and did not include the entire metropolitan area.

## IV. **Methodology:**

For this project, a python notebook was created in the IBM Skills Network website.

The analysis includes two distinct parts to meet the assignment requirements: first, a k means clustering of healthcare venues with the Foursquare API and second, an analysis of discharge and healthcare charges based on healthcare census data from the State of Washington Department of Health (DOH).

### K-means clustering:

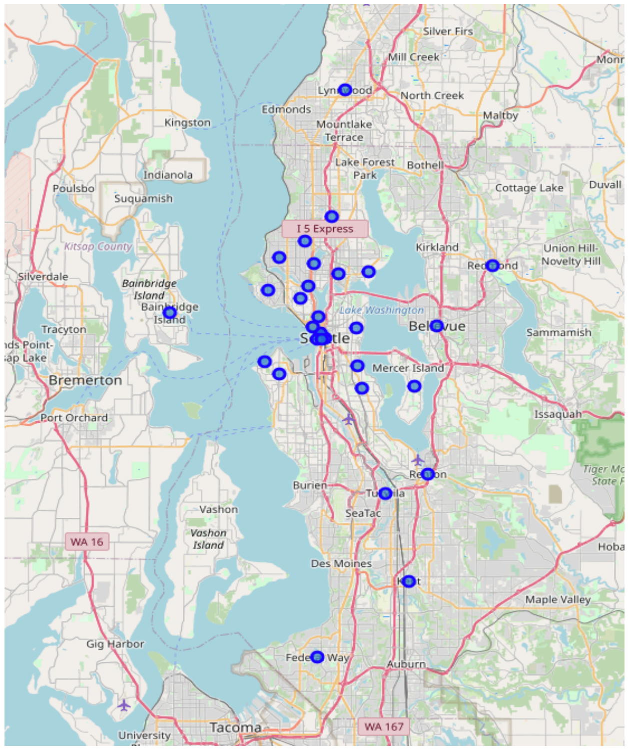
In the first part, the Folium Python package was utilized in conjunctions with Foursquare API data to screen venues which specifically met the search criterial "healthcare service".   First, City of Seattle neighborhood data had to be scrapped into a workable dataframe.   The useful site information was in a PDF format, and although there are some packages in Python which allow pdf format files to be scraped, often these packages remain difficult and cumbersome.

PDF formats were conceived as a method to ensure format consistency across various and numerous computing systems.   As such, the PDF is not in a file format such as markup languages, CSV files, spreadsheet formats, or word processing applications.    Given time constraints for this project, instead this data was placed into an MS Excel document and then read using the CSV library "reader" method.

A Foursquare Client connection had to be established via the Foursquare API.  Based on this API, a Python loop was coded to pull the latitudes and longitudes of the various Seattle Neighborhood names directly and pull these into a single dataframe:

| | Zip Code | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | 98003 | Federal Way, WA | 47.313494 | -122.339310 |
| 1 | 98005 | Bellevue, WA | 47.614422 | -122.192337 |
| 2 | 98037 | Lynnwood, WA | 47.827866 | -122.305393 |
| 3 | 98040 | Mercer Island, WA | 47.560207 | -122.220142 |
| 4 | 98052 | Redmond, WA | 47.669414 | -122.123877 |
| 5 | 98055 | Renton, WA | 47.479908 | -122.203450 |
| 6 | 98101 | Seattle, WA | 47.603832 | -122.330062 |
| 7 | 98101 | Downtown Seattle, WA | 47.604872 | -122.333458 |
| 8 | 98102 | Capital Hill Seattle, WA | 47.608263 | -122.335190 |
| 9 | 98103 | Greenwood, WA | 47.690981 | -122.354877 |

Based on this pandas dataframe, the Folium library can be used to plot the select neighborhoods of Seattle:

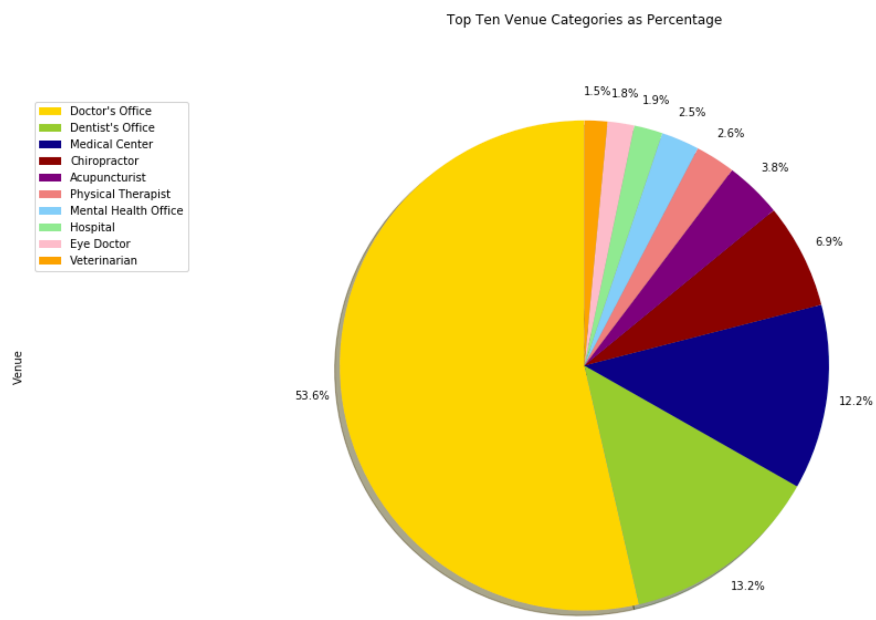



Next, the analysis focused specifically on health services for each of the neighborhoods. Initially, an attempt of using the Foursquare API query around "health services" was used. Starting with Federal Way (this was the first neighborhood in the Seattle Region that was in the dataframe), along with using a 1500 meter (approximately 1mi radius) and a limit of 100 results, the result was too limited. Foursquare indicated that a broader set of filter criteria would be needed to be used to generate a result:


```
{'meta': {'code': 200, 'requestId': '5ef78611546946258f87f366'},
 'response': {'warning': {'text': 'There aren\'t a lot of results for "health services." Try something more general, reset your filters, or expand the search area.'},
  'headerLocation': 'Federal Way',
  'headerFullLocation': 'Federal Way',
  'headerLocationGranularity': 'city',
  'query': 'health services',
  'totalResults': 0,
  'suggestedBounds': {'ne': {'lat': 47.32699401350001,
    'lng': -122.31943555901097},
   'sw': {'lat': 47.299993986499985, 'lng': -122.35918504098902}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': []}]}}
```


Refining the analysis to include a 3000m search radius, coupled with a single keyword of "healthcare" generated the full limit of 100 venues meeting the criteria.

For the 30 total Seattle neighborhoods identified in the analysis, a healthcare venues dataframe was generated. These venues covered a total of 29 unique venue categories, but many of these were of incredibly small sizes. Doctor and dentist offices were vastly more common. For the top ten venue categories, a pie chart categorization was generated




To better categorize the healthcare venues by neighborhoods, the analysis was further refined using a k-means clustering approach which required that the full dataframe be encoded using the one-hot encoding approach which provided dummy-variables for each of the venue categories.

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Alki Beach, WA | Alternative Healer | Doctor's Office | Veterinarian | Hospital Ward | Assisted Living | Building | Chiropractor | Daycare | Dentist's Office | Emergency Room |
| **1** | Bainbridge Island, WA | Doctor's Office | Dentist's Office | Acupuncturist | Chiropractor | Medical Center | Hospital Ward | Alternative Healer | Assisted Living | Building | Daycare |
| **2** | Ballard, WA | Dentist's Office | Massage Studio | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Doctor's Office |
| **3** | Bellevue, WA | Doctor's Office | Dentist's Office | Chiropractor | Medical Center | Physical Therapist | Veterinarian | Hospital | Urgent Care Center | Medical Lab | Mental Health Office |
| **4** | Belltown, WA | Doctor's Office | Dentist's Office | Chiropractor | Medical Center | Mental Health Office | Eye Doctor | Veterinarian | Alternative Healer | Acupuncturist | Marijuana Dispensary |
| **5** | Capital Hill Seattle, WA | Doctor's Office | Medical Center | Optical Shop | Veterinarian | Hospital | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare |
| **6** | Columbia City, WA | Dentist's Office | Acupuncturist | Mental Health Office | Doctor's Office | Maternity Clinic | Veterinarian | Office | Financial or Legal Service | Alternative Healer | Assisted Living |
| **7** | Downtown Seattle, WA | Doctor's Office | Medical Center | Office | Veterinarian | Hospital | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare |
| **8** | Federal Way, WA | Dentist's Office | Doctor's Office | Acupuncturist | Chiropractor | Daycare | Medical Center | Eye Doctor | Hospital Ward | Alternative Healer | Assisted Living |
| **9** | Fremont Seattle, WA | Doctor's Office | Acupuncturist | Medical Center | Chiropractor | Physical Therapist | Dentist's Office | Mental Health Office | Health & Beauty Service | Alternative Healer | Marijuana Dispensary |
| **10** | Greenlake, WA | Physical Therapist | Hospital | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Dentist's Office |
| **11** | Greenwood, WA | Doctor's Office | Medical Center | Physical Therapist | Dentist's Office | Veterinarian | Chiropractor | Building | Massage Studio | Maternity Clinic | Acupuncturist |
| **12** | Kent, WA | Dentist's Office | Doctor's Office | Medical Center | Laboratory | Chiropractor | Hospital Ward | Alternative Healer | Assisted Living | Building | Daycare |
| **13** | Laurelhurst, WA | Hospital | Doctor's Office | Medical Center | Hospital Ward | Laboratory | Medical Lab | Emergency Room | Eye Doctor | Alternative Healer | Assisted Living |
| **14** | Lynnwood, WA | Dentist's Office | Veterinarian | Doctor's Office | Chiropractor | Acupuncturist | Optical Shop | Office | Alternative Healer | Assisted Living | Building |
| **15** | Madrona, WA | Dentist's Office | Acupuncturist | Alternative Healer | Mental Health Office | Hospital Ward | Assisted Living | Building | Chiropractor | Daycare | Doctor's Office |
| **16** | Magnolia Seattle, WA | Dentist's Office | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Doctor's Office | Emergency Room |
| **17** | Mercer Island, WA | Dentist's Office | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Doctor's Office | Emergency Room |
| **18** | Mount Baker Seattle, WA | Physical Therapist | Dentist's Office | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Doctor's Office |
| **19** | Northgate, WA | Dentist's Office | Chiropractor | Medical Center | Mental Health Office | Doctor's Office | Veterinarian | Alternative Healer | Assisted Living | Acupuncturist | Medical Lab |
| **20** | Pioneer Square, WA | Doctor's Office | Chiropractor | Dentist's Office | Medical Center | Mental Health Office | Physical Therapist | Eye Doctor | Acupuncturist | Health & Beauty Service | Medical Lab |
| **21** | Queen Anne, WA | Doctor's Office | Dentist's Office | Chiropractor | Acupuncturist | Physical Therapist | Eye Doctor | Hospital Ward | Alternative Healer | Assisted Living | Building |
| **22** | Redmond, WA | Dentist's Office | Physical Therapist | Doctor's Office | Veterinarian | Medical Center | Chiropractor | Acupuncturist | Eye Doctor | Optical Shop | Hospital |
| **23** | Renton, WA | Dentist's Office | Chiropractor | Mental Health Office | Medical Lab | Medical Center | Veterinarian | Alternative Healer | Doctor's Office | Hospital | Assisted Living |
| **24** | Seattle Chinatown - international District, WA | Doctor's Office | Chiropractor | Dentist's Office | Mental Health Office | Medical Center | Veterinarian | Hospital | Alternative Healer | Assisted Living | Building |
| **25** | Seattle, WA | Doctor's Office | Medical Center | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Dentist's Office |
| **26** | South Lake Union, WA | Doctor's Office | Dentist's Office | Chiropractor | Medical Lab | Medical Center | Eye Doctor | Veterinarian | Hospital | Office | Alternative Healer |
| **27** | Tukwila, WA | Doctor's Office | Dentist's Office | Chiropractor | Mental Health Office | Eye Doctor | Urgent Care Center | Veterinarian | Hospital | Alternative Healer | Assisted Living |
| **28** | University District, WA | Doctor's Office | Dentist's Office | Medical Center | Chiropractor | Acupuncturist | Physical Therapist | Hospital | Marijuana Dispensary | Medical Lab | Mental Health Office |
| **29** | West Seattle, WA | Doctor's Office | Dentist's Office | Acupuncturist | Medical Center | Chiropractor | Medical Lab | Alternative Healer | Physical Therapist | Hospital | Assisted Living |

With the top venue categories for each Neighborhood developed, the k-means clustering approach was applied to these venues. As a an unstructured (e.g. “unlabeled”) clustering technique, k-means assigns the clusters based upon Minkoskian distances calculated amongst the data points. Ultimately, the neighborhoods were clustered using a k=5, and the Folium package was utilized to map these calculated clusters.

Washington Department of Health Data:

The second portion of the analysis scraped data from an Excel spreadsheet which utilized 2019 Hospital Census Charge data separated by payor. The pandas library was used to read in the data which then had to be appropriately scrubbed given a large number NaN values given formatting in the Excel workbook.

The census data included information and ratios based upon several key variables including:

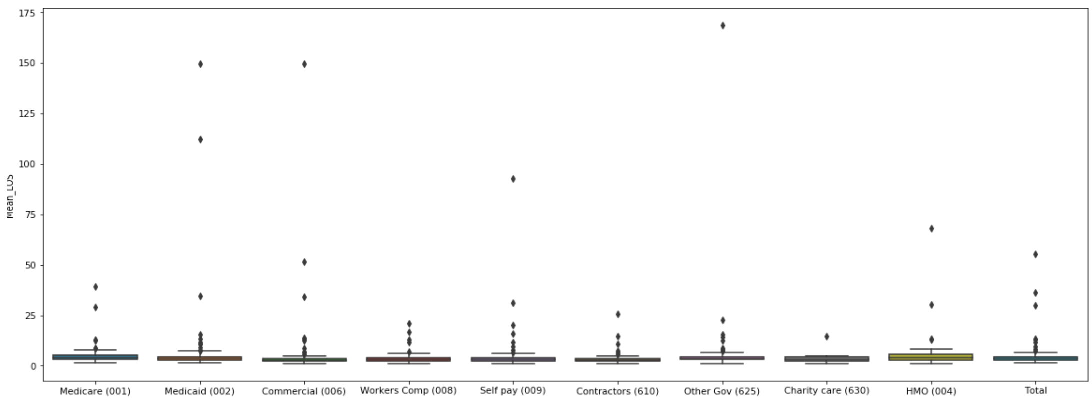
| **Fields Name** | **Description** |
|---|---|
| Hospital_name | Name of the facility |
| Type | Type of facility – Includes Acute care, urgent care, skilled nursing facility, etc. |
| Lic | State Healthcare License Number |
| Payername | |
| Discharges | Total number of discharged patients for period |
| Total_days | Pateint days in hospital as defined under the two midnight rule |
| Total_charges | Total amount charged for the care received |
| Mean_LOS | Length of stay, rounded to nearest integer value. The Length of stay is the total number of days divided by the total number of discharges |
| Mean_charge_per_discharge | Total charges divided by the number of discharges |
| Mean_Charge_per_Day | Total charges divided the number of patient days |
| Casemix_index | An adjustment factor based on the relative population sickness being treated at the facility. Charge metrics are divided by this factor. If the factor is less than one the Charge metric will increase, if the factor is greater than one the metric will decrease. |
| Casemix_adjusted Charge per Disch | Calculated charge based on casemix index as described above. |

Once the dataframe was suitably clean, it was possible to consider how to use the data to make judgements about desirable hospitals in Seattle. This particular set included various care facilities, not just hospitals, but also swing bed units, urgent care clinics, doctors, dentists. The example data frame also included multiple lines per facility (see below).

| | Hospital_name | Type | Lic | payername | Discharges | Total_days | Total_charges | Mean_LOS | Mean_Charge_per_Discharge | Mean_Charge_per_Day | Casemix_per_Discharge | Casemix_Index | Count_of_Weights | Sum_of_Weights |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Arbor Health | Acute Care | 173 | Medicare (001) | 146 | 479 | 4018452.49 | 3.28 | 27523.65 | 8389.25 | 42089.36 | 0.65 | 143 | 93.51 |
| 1 | Arbor Health | Acute Care | 173 | Medicaid (002) | 17 | 43 | 568882.73 | 2.53 | 33463.69 | 13229.83 | 52435.25 | 0.64 | 16 | 10.21 |
| 2 | Arbor Health | Acute Care | 173 | Commercial (006) | 22 | 64 | 735537.09 | 2.91 | 33433.50 | 11492.77 | 42205.04 | 0.79 | 22 | 17.43 |
| 3 | Arbor Health | Acute Care | 173 | Workers Comp (008) | 1 | 4 | 35551.30 | 4.00 | 35551.30 | 8887.83 | 71043.21 | 0.50 | 1 | 0.50 |
| 4 | Arbor Health | Acute Care | 173 | Self pay (009) | 4 | 8 | 90801.10 | 2.00 | 22700.28 | 11350.14 | 34393.42 | 0.66 | 4 | 2.64 |
| 5 | Arbor Health | Swing Bed Unit | 173S | Medicare (001) | 89 | 1244 | 2915834.50 | 13.98 | 32762.19 | 2343.92 | 48389.06 | 0.68 | 88 | 59.58 |
| 6 | Arbor Health | Swing Bed Unit | 173S | Medicaid (002) | 5 | 3560 | 1523715.71 | 712.00 | 304743.14 | 428.01 | 530864.64 | 0.57 | 5 | 2.87 |
| 7 | Arbor Health | Swing Bed Unit | 173S | Commercial (006) | 1 | 13 | 25220.60 | 13.00 | 25220.60 | 1940.05 | 22696.96 | 1.11 | 1 | 1.11 |
| 8 | Arbor Health | Swing Bed Unit | 173S | Workers Comp (008) | 2 | 27 | 52786.00 | 13.50 | 26393.00 | 1955.04 | 46368.72 | 0.57 | 2 | 1.14 |
| 9 | Astria Regional Medical Center | Acute Care | 102 | Medicare (001) | 1269 | 5134 | 95660509.25 | 4.05 | 75382.59 | 18632.74 | 71749.72 | 1.05 | 1256 | 1319.59 |
| 10 | Astria Regional Medical Center | Acute Care | 102 | Medicaid (002) | 617 | 2545 | 42391689.37 | 4.12 | 68706.14 | 16656.85 | 71684.69 | 0.96 | 607 | 581.78 |
| 11 | Astria Regional Medical Center | Acute Care | 102 | Commercial (006) | 260 | 884 | 22129592.54 | 3.40 | 85113.82 | 25033.48 | 69615.54 | 1.22 | 260 | 317.88 |
| 12 | Astria Regional Medical Center | Acute Care | 102 | Workers Comp (008) | 12 | 49 | 1160059.55 | 4.08 | 96671.63 | 23674.68 | 80198.04 | 1.21 | 12 | 14.46 |
| 13 | Astria Regional Medical Center | Acute Care | 102 | Self pay (009) | 49 | 140 | 2872286.32 | 2.86 | 58618.09 | 20516.33 | 63552.68 | 0.92 | 49 | 45.20 |
| 14 | Astria Regional Medical Center | Acute Care | 102 | Contractors (610) | 1 | 5 | 672972.30 | 5.00 | 672972.30 | 134594.46 | 426520.35 | 1.58 | 1 | 1.58 |
| 15 | Astria Regional Medical Center | Acute Care | 102 | Other Gov (625) | 68 | 254 | 5871325.25 | 3.74 | 86343.02 | 23115.45 | 73523.93 | 1.17 | 67 | 78.68 |
| 16 | Astria Regional Medical Center | Acute Care | 102 | Charity care (630) | 5 | 15 | 395325.83 | 3.00 | 79065.17 | 26355.06 | 66329.45 | 1.19 | 5 | 5.96 |
| 17 | Astria Sunnyside Hospital | Acute Care | 198 | Medicare (001) | 454 | 2105 | 23582316.31 | 4.64 | 51943.43 | 11203.00 | 58570.44 | 0.89 | 449 | 398.20 |
| 18 | Astria Sunnyside Hospital | Acute Care | 198 | Medicaid (002) | 795 | 1708 | 15967633.26 | 2.15 | 20085.07 | 9348.73 | 53391.78 | 0.38 | 788 | 296.43 |
| 19 | Astria Sunnyside Hospital | Acute Care | 198 | Commercial (006) | 173 | 411 | 5032563.97 | 2.38 | 29089.97 | 12244.68 | 61557.50 | 0.47 | 172 | 81.28 |
| 20 | Astria Sunnyside Hospital | Acute Care | 198 | Workers Comp (008) | 4 | 16 | 149922.14 | 4.00 | 37480.54 | 9370.13 | 54232.97 | 0.69 | 4 | 2.76 |
| 21 | Astria Sunnyside Hospital | Acute Care | 198 | Self pay (009) | 24 | 68 | 1332895.83 | 2.83 | 55537.33 | 19601.41 | 68953.98 | 0.81 | 24 | 19.33 |
| 22 | Astria Sunnyside Hospital | Acute Care | 198 | Other Gov (625) | 11 | 31 | 985504.84 | 2.82 | 89591.35 | 31790.48 | 91845.36 | 0.98 | 11 | 10.73 |
| 23 | Astria Sunnyside Hospital | Swing Bed Unit | 198S | Medicare (001) | 2 | 21 | 68504.13 | 10.50 | 34252.07 | 3262.10 | 59724.84 | 0.57 | 2 | 1.15 |

For purposes of this analysis, analysis of length of stay and an analysis of individual hospital charges was conducted.

First, the analysis considers the overall length of stay (LOS) of patients relative to their payer classification (i.e. payername). Looking at this element can be used to determine whether certain payer classifications have shorter or longer overall lengths of stay associated with them.



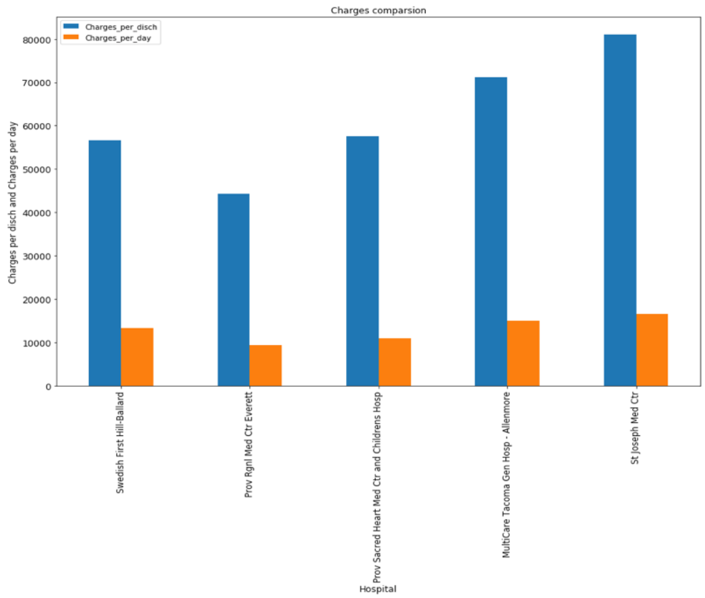

Notice that the scale of the boxplot graphs when considering Mean Length of Stay relative to the payer type causes the graph to become highly compressed and difficult to read. This is due to the relatively large number of outliers in the dataset. Certain payers have enough instances outside of the interquartile range, that to graph all data points, the boxplots themselves become unreadable. Despite this data trend, generally the length of stay will remain in the 2 to 6-day range. We can filter the data to exclude mean LOS beyond 10 days, for example to get a more discerning set of resulting boxplots which when compared give a visual representation of LOS sensitivity to payer class.

The second part of the analysis focuses on Charges and considers only acute care hospital settings for the analysis. It was first necessary to use pandas groupby methods to screen out this "type" of facility. Additionally, the groupby method was employed to consolidate all payer classes, as we want to compare Charges per discharge and Charges per day for the entire population of acute care facilities.

Initially, the screening methodology was intended to look at those acute care sites that had the lowest overall Charges per discharge and Charge per day, as shown in the following table.

| Hospital_name | Discharges | Total_days | Total_charges | Charges_per_disch | Charges_per_day |
|---|---|---|---|---|---|
| East Adams Rural Healthcare | 6 | 8 | 26492.00 | 4415.33 | 3311.50 |
| Garfield County Public Hospital District | 14 | 88 | 70473.00 | 5033.79 | 800.83 |
| Odessa Memorial Healthcare Center | 38 | 98 | 268151.98 | 7056.63 | 2736.24 |
| Columbia Basin Hosp | 260 | 732 | 1943316.14 | 7474.29 | 2654.80 |
| Cascade Med Ctr | 116 | 306 | 1052800.58 | 9075.87 | 3440.52 |

This approach, however, proved to be skewed as the results were predominantly for facilities with very low overall discharges, especially when compared to the larger facilities.

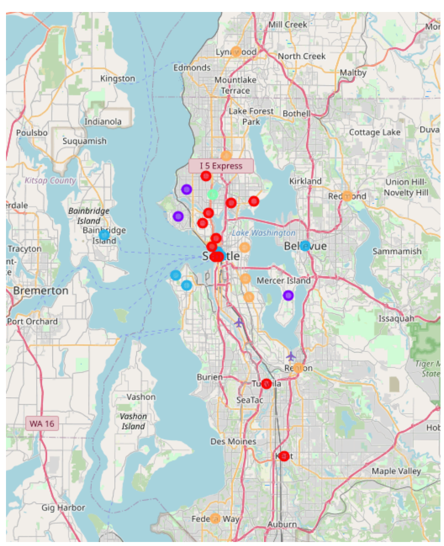



Instead, the analysis was modified to consider the top five facilities by discharge volume as the basis of comparison. This provides a fairer comparison, particularly as these facilities tend to be the largest in Seattle city proper.

## V. <u>Results and Discussion:</u>

**(Please note that for explanatory purposes, I found it more logical to incorporate the results and discussion components of the project together. I believe that doing so helps present a stronger overall narrative than separating the two.)**

<u>K-means cluster results:</u>
Based upon the k means clustering the following 5 clusters were generated for the Seattle healthcare venues.



The bulk of the Cluster 1 is around the central business area and includes Downtown Seattle, Pioneer Square and the International District as noted by the red dots. The predominant defining elements which differentiate the clusters are the prominence of particular common venues which are not overall common to the others.

This becomes more apparent when particularly notable when we look at the comparison breakdown between clusters. Take, for example Clusters 1, 2, and 4:

Cluster 1:

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | Greenwood, WA | Doctor's Office | Medical Center | Physical Therapist | Dentist's Office | Veterinarian | Chiropractor | Building | Massage Studio | Maternity Clinic | Acupuncturist |
| 10 | Fremont Seattle, WA | Doctor's Office | Acupuncturist | Medical Center | Chiropractor | Physical Therapist | Dentist's Office | Mental Health Office | Health & Beauty Service | Alternative Healer | Marijuana Dispensary |
| 12 | Seattle Chinatown - international District, WA | Doctor's Office | Chiropractor | Dentist's Office | Mental Health Office | Medical Center | Veterinarian | Hospital | Alternative Healer | Assisted Living | Building |
| 13 | Pioneer Square, WA | Doctor's Office | Chiropractor | Dentist's Office | Medical Center | Mental Health Office | Physical Therapist | Eye Doctor | Acupuncturist | Health & Beauty Service | Medical Lab |
| 14 | University District, WA | Doctor's Office | Dentist's Office | Medical Center | Chiropractor | Acupuncturist | Physical Therapist | Hospital | Marijuana Dispensary | Medical Lab | Mental Health Office |
| 15 | Laurelhurst, WA | Hospital | Doctor's Office | Medical Center | Hospital Ward | Laboratory | Medical Lab | Emergency Room | Eye Doctor | Alternative Healer | Assisted Living |
| 17 | South Lake Union, WA | Doctor's Office | Dentist's Office | Chiropractor | Medical Lab | Medical Center | Eye Doctor | Veterinarian | Hospital | Office | Alternative Healer |
| 18 | Queen Anne, WA | Doctor's Office | Dentist's Office | Chiropractor | Acupuncturist | Physical Therapist | Eye Doctor | Hospital Ward | Alternative Healer | Assisted Living | Building |
| 24 | Belltown, WA | Doctor's Office | Dentist's Office | Chiropractor | Medical Center | Mental Health Office | Eye Doctor | Veterinarian | Alternative Healer | Acupuncturist | Marijuana Dispensary |
| 28 | Kent, WA | Dentist's Office | Doctor's Office | Medical Center | Laboratory | Chiropractor | Hospital Ward | Alternative Healer | Assisted Living | Building | Daycare |
| 29 | Tukwila, WA | Doctor's Office | Dentist's Office | Chiropractor | Mental Health Office | Eye Doctor | Urgent Care Center | Veterinarian | Hospital | Alternative Healer | Assisted Living |

Cluster 2:

```
[38]: seattle_merged.loc[seattle_merged['Cluster Labels'] == 1, seattle_merged.columns[[1] + list(range(5, seattle_merged.shape[1]))]]
```

[38]:

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Mercer Island, WA | Dentist's Office | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Doctor's Office | Emergency Room |
| 16 | Ballard, WA | Dentist's Office | Massage Studio | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Doctor's Office |
| 27 | Magnolia Seattle, WA | Dentist's Office | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Doctor's Office | Emergency Room |

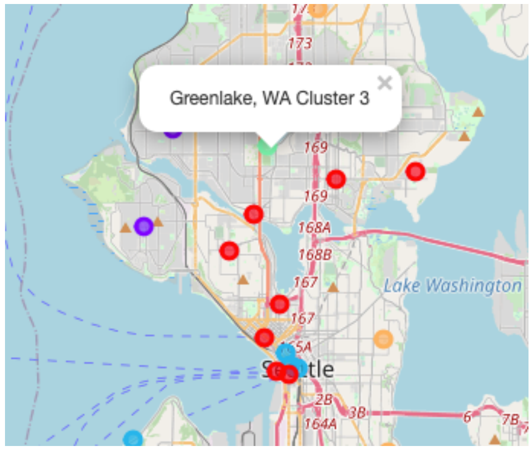
Cluster 4:

```
seattle_merged.loc[seattle_merged['Cluster Labels'] == 4, seattle_merged.columns[[1] + list(range(5, seattle_merged.shape[1]))]]
```

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | Greenlake, WA | Physical Therapist | Hospital | Veterinarian | Hospital Ward | Alternative Healer | Assisted Living | Building | Chiropractor | Daycare | Dentist's Office |

The most common venues in Cluster 1 were "Doctor's Office" and Dentist's office", following closely by "Chiropractor". Which is sensible as the initial data analysis and pie chart completed indicates that the vast bulk of healthcare venues were of these categories. Curiously, Cluster 2 includes few Doctor's offices and also includes "Hospital Ward" as a common distinction from Cluster 1. In the 3rd and 4th most common venue.

Cluster 4 is quite unique in that it only includes a single Seattle neighborhood – Greenlake, where the 1st most common venue is "Physical therapist".

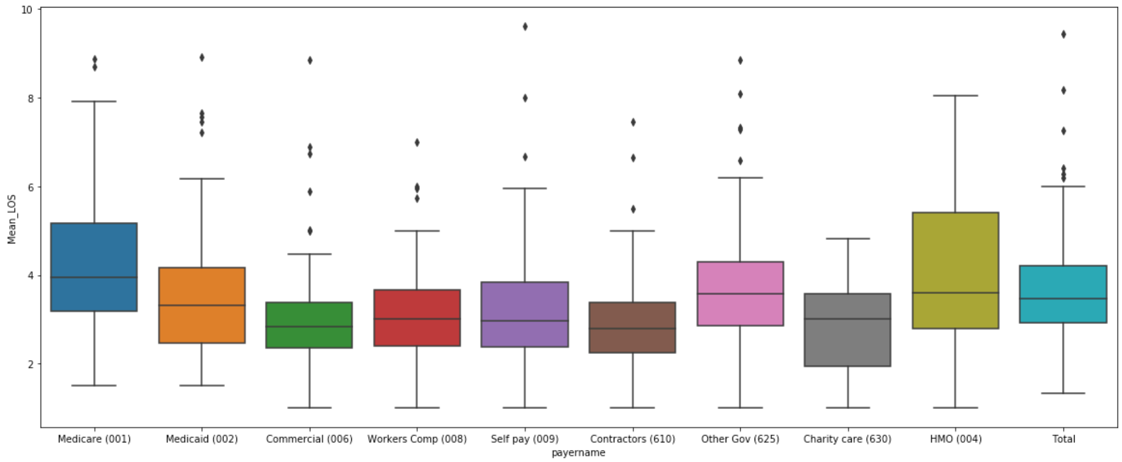



It seems that based on the k – means structuring, the definitional categories applied by Foursquare may lack the necessary specificity required to draw consistent and meaningful conclusions. It is curious that the keyword query "healthcare" also includes venues categorized as "daycare" or simply as "building". Additionally, there are categories such as "Hospital Ward" which differs from "Hospital" which again is differentiable by "Medical Center". Further refinement and understanding of Foursquare definitions would need to be researched to be able to make better conclusions and groupings using the k means algorithm.

Simultaneously, the choice of combination of filter keywords used to identify the types of healthcare delivery locations often results in too narrow a search in many cases. To be of greater use and benefit, Foursquare's algorithms itself may need to be redesigned or reproached when it comes to healthcare. Obviously Foursquare's audience is far broader and its intent is not specific to the healthcare industry.

<u>Washington Department of Health Data:</u>

With consideration of only lengths of stay less than 10 days, the overall

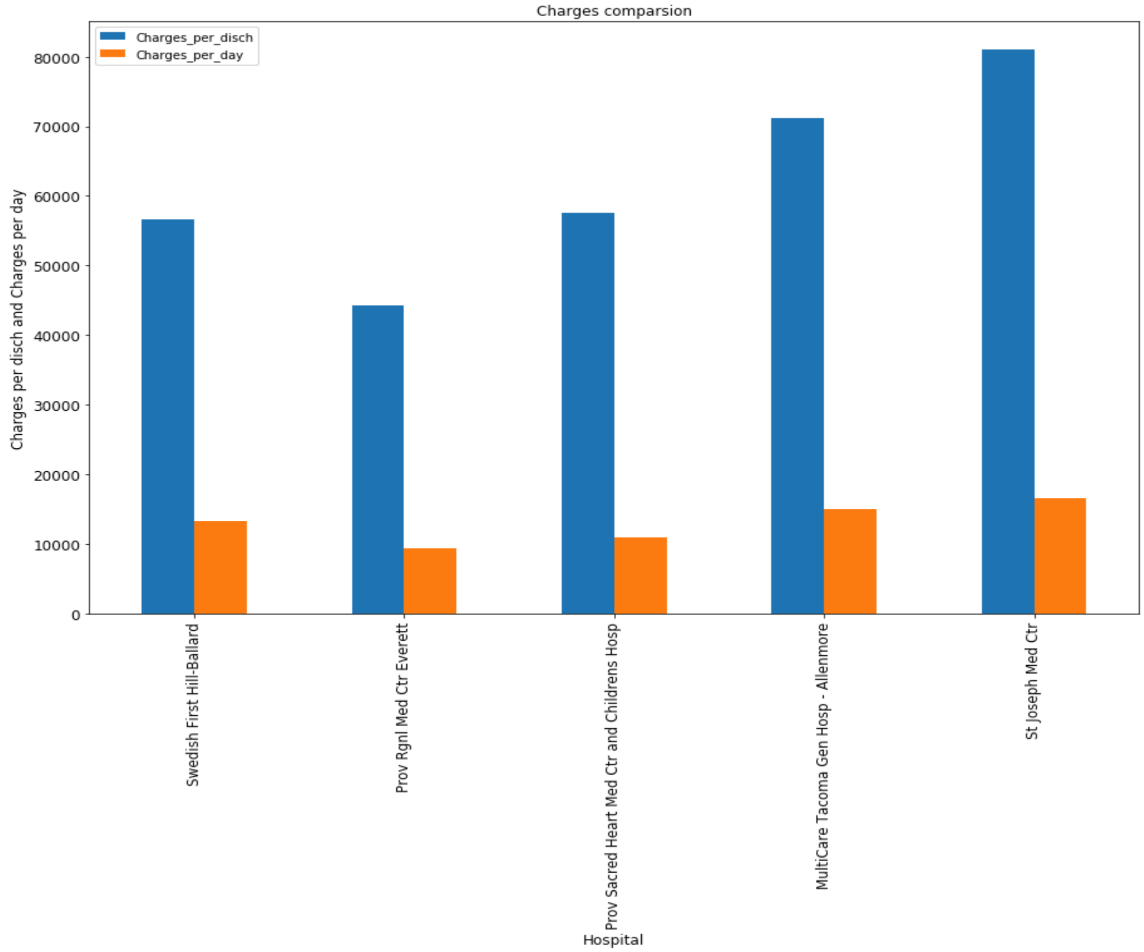



It is interesting to consider the relative sizes of the interquartile range of the various boxplots. Note that Commercial payer type has much narrower overall bounds that Medicare or HMO payer types for example. This could suggest that Commercial and Contractor payer is overall far better at controlling length of stay, as they face greater financial incentive to reduce length of stay as opposed to government payer types.

For the last segment of this analysis, the Charge data for the largest hospital for discharge volume was considered.

| Hospital_name | Discharges | Total_days | Total_charges | Charges_per_disch | Charges_per_day |
|---|---|---|---|---|---|
| Swedish First Hill-Ballard | 79548 | 340442 | 4509615860.90 | 56690.50 | 13246.36 |
| Prov Rgnl Med Ctr Everett | 67052 | 315352 | 2965003123.14 | 44219.46 | 9402.20 |
| Prov Sacred Heart Med Ctr and Childrens Hosp | 62680 | 330946 | 3605333668.00 | 57519.68 | 10894.02 |
| MultiCare Tacoma Gen Hosp - Allenmore | 47902 | 227652 | 3411209439.04 | 71212.26 | 14984.32 |
| St Joseph Med Ctr | 46938 | 229790 | 3802601730.90 | 81013.29 | 16548.16 |

This tabular format corresponds to the following bar chart:




Overall, it appears that Providence Sacred Heart balances charges per discharge and charges per day the best of these top 5 hospitals by discharge volume. Swedish First Hill hospital in Ballard has substantially more discharges (nearly 20% more) than any of the other facilities measured here. Simultaneously, both its charges per discharge and charges per day are commiserate with the other top-five volume hospitals.

In the future, a more comprehensive statistical test could be used to determine if these hospitals are truly different from one another in terms of charges performance. For example, a t-test could be used based on sample data of these hospitals. Such an analysis is not possible here because the State of Washington Department of Health data provided is only in summary format. The actual population of charge data would need to be determined to perform a more robust analysis.

## VI. **Conclusion:**

This project demonstrates how useful Python as a tool in Data Science can be in drawing out basic understanding from fairly large and desperate data sets. This particular project has looked at one particular American city – Seattle, Washington - to understand some key facets of their healthcare marketplace. Healthcare is surprisingly different region to region and city to city due to differences in health policy, overall investment, and health resource bases. Expanding upon the concepts discussed here, it may be useful to consider other major metropolitan areas when looking at clustering health resources.

Due to time constraints, and that this Capstone project is serving only as an initial foray into data science concepts, these more robust metrics will not be explored further here.