**INDUSTRY CLUSTER ANALYSIS FOR OHIO STATE**

**AND ADAMS COUNTY, OHIO**

by

**Ateet Maharjan**

[www.linkedin.com/in/ateetmaharjan](http://www.linkedin.com/in/ateetmaharjan)

To view the codes, please check the github profile: <https://github.com/ateetmaharjan/Webscraping-for-Industry-Cluster>

**TABLE OF CONTENTS**

[INTRODUCTION 1](#_Toc39830753)

[1.1. Background 1](#_Toc39830754)

[1.2. Scope of work 2](#_Toc39830755)

[methods 2](#_Toc39830756)

[2.1 Data 2](#_Toc39830757)

[2.2 Dataset preparation 2](#_Toc39830758)

[ANALYSIS AND interpretation OF findings 3](#_Toc39830759)

[3.1 Industry Cluster Analysis for Ohio 3](#_Toc39830760)

[3.2 Shif-Share Analysis for Ohio 5](#_Toc39830761)

[3.3 Industry Cluster Analysis for Adams County, Ohio 5](#_Toc39830762)

[DISCUSSION ON CHALLENGES 8](#_Toc39830763)

[REFERENCES 9](#_Toc39830764)

[APPENDIX 10](#_Toc39830765)

# INTRODUCTION

## Background

Industry clusters are a framework for economic development that views the regional

economy as a collection of clusters, rather than many individual firms influenced by specific

events. This study analyzes twenty industry clusters at Ohio State along with Adams County in two ways: using location quotients and using shift-share analysis. Location quotients and their changes help identify strengths and weaknesses in the region’s economy. Industry clusters that represent strengths in the region are those in which the region specializes and its concentration in the region is increasing. The shift-share analysis helps to identify the impact of region-specific factors on regional economic growth. For the clusters in which the region has a positive impact, the region has a competitive advantage.

An LQ compares the fraction of the region’s employment in a particular industry cluster to the fraction of the nation’s employment in the same industry cluster.

* LQ = 1 indicates that the region employs the same fraction of its workforce in the industry cluster as does the nation as a whole.
* LQ > 1 indicates that the region employs a larger fraction of its workforce in the industry cluster (less than one a smaller fraction) than does the nation. When the LQ exceeds 1.0, the region “specializes” in the industry cluster. The region has a concentration in the industry cluster.

Industry clusters are often displayed in a three-dimensional (LQ, percentage change in LQ, and cluster employment) bubble chart. The size of bubble represents industry cluster’s employment. Larger the cluster, larger is the industry cluster’s employment. The four quadrants include:

* Upper right: Star clusters
  + Regional employment fraction > nation, employment increased over a period
  + specialized as compared to the nation and becoming more specialized.
* Upper left: Mature clusters
  + Regional employment fraction > nation, employment decreased over a period
  + specialized as compared to the nation and becoming less specialized.
* Lower right: Emerging clusters
  + Current fractional employment fraction < nation, employment increased over a period
  + Less specialized as compared to the nation and may become specialized in future
* Lower left: Transforming clusters
  + Regional employment fraction < nation, employment decreased over a period
  + Less specialized and unlikely to become specialized.

## Scope of work

The data task requires the creation of industry cluster bubble charts, which are a common economic development research technique used to identify industries in which a region has a comparative advantage. An industry cluster bubble chart needs to be prepared for the State of Ohio, and then a separate chart for Adams County, Ohio. Specifically, there are three components of this data task:

* First, familiarize with the methods, as well as identify the data to use.
* Next, create the dataset needed to produce the charts.
* Then, create the traditional industry cluster bubble charts for Ohio and Adams County, and also a different chart to illustrate the same concept in own way to demonstrate alternative data visualization techniques.

# methods

To create the industry cluster bubble chart, data went through series of cleaning and preparation for the format to build a figure. The methods followed are as follows:

## Data

The data for preparing cluster chart was obtained from the Quarterly Census of Employment and Wages- Bureau of Labor Statistics. The data used for analysis consisted of annual average of employment, wage, and location quotient column for private sector industry clusters. The industry clusters were based on the North American Industry Classification System (NAICS) codes. For the analysis, 20 industry clusters (NAICS sectors) were used for Ohio and US, whereas only 15 industry clusters were used for Adams county. The data for Adams country excluded rows with suppressed employment and wages. Therefore only 15 clusters were used for analysis. Two years of data: 2015 and 2018 was used for the industry cluster analysis.

The data on the website was scraped using Python, looping over the required pages. The required information from the page was inspected using developer options in the browser. Modules: “beautiful soup” and “selenium” were installed and used for scraping data from the website’s API (application programming interface).

## Dataset preparation

The data was handled using excel. The scraped data were clean enough. Excel was used to calculate required values for Location Quotient (LQ), % change in LQ from 2015 -2018 and % change in employment from 2015-2018. Also, other numerical values were calculated using the formula below:

*Location Quotient* = [Eri/Er] / [Eni/En]

i.e. proportion of employment of industry cluster for region divided by nation

where,

Eri = Employment of Industry Cluster in Region

Er = Employment of Region

Eni = Employment of industry cluster in Nation

En = Employment of Nation

Along with LQ, other values were also calculated other variables for share-shift analysis for Ohio State.

* *National Share* = % change in En \* Eri (year 2015)
* *Industry Cluster Share* = (%change in Eni - %change in En) \* Eri (2015-initial year)
* *Regional Share*  = Actual change in Eri - [NS+IS]

After creating the required variables, data were imported in Python, and after performing some basic data handling, the industry cluster bubble chart was prepared.

# ANALYSIS AND interpretation OF findings

## Industry Cluster Analysis for Ohio

|  |
| --- |
| A picture containing large  Description automatically generated |

Fig 1. Industry Cluster Bubble Chart for the region - Ohio State, US

Figure 1 shows industry cluster analysis, including all industry sectors. The inclusion of unclassified industry cluster made it difficult for clear interpretation. It does not clearly reveal the other industry clusters in the region, so the data was excluded from the visualization in Figure 2.

Figure 2 shows the bubble chart for the industry clusters in Ohio states with unclassified excluded. The upper right quadrant includes star clusters such as Manufacturing, Wholesale Trade, Transportation, and Warehousing. In these clusters, the fraction of employment in the region is larger than that in the nation and has increased over the 3 years relative to the nation. They are specialized as compared to the nation and are becoming more specialized. Ohio’s fraction of employment in Finance and Insurance cluster increased substantially more than in the nation compared with many of the other industry clusters in the region. The cluster is becoming even more specialized. Manufacturing has the highest LQ value in this cluster, and its degree of concentration increased from 2015-2018 compared to the national level. Also, it is the largest cluster in this category with 698,950 employees in 2018.

|  |
| --- |
| A close up of a map  Description automatically generated |

Fig 2. Industry Cluster Bubble Chart for the region - Ohio State, US (removed extremes)

The mature cluster includes “Management of companies and enterprises” and “healthcare assistance.” Ohio is specialized in these clusters; however, they require some attention, since they are becoming less specialized. Management of companies and enterprises cluster has the highest LQ of 1.62 in the whole Ohio states. The fraction of employment for this cluster in Ohio is higher compared to the national level in 2018; however, employment has decreased over three years. The industry cluster healthcare has the highest employment workforce in 2018 (778907 jobs); however, the cluster experienced a decline (3.81 % growth) in employment compared to national level (7.46% growth) from 2015 – 2018.

Ohio state has important emerging clusters, including “Agriculture, forestry, fishing and hunting” and “Utilities.” Ohio is not specialized in these regions; however, these clusters hold some potential for the future. The industry sector “Agriculture” has increased the workforce (4.66 %) over the three years period compared to the national average of 1.15 %.

Similarly, transforming clusters include “Mining,” “Information,” “Real state,” “Administrative,” “Educational services,” and “Arts, entertainment.” They are less specialized and unlikely to become specialized. However, there are some clusters with a workforce equal to the national level in 2018. ﻿While transforming clusters hold less potential for Ohio state as a whole, some of these clusters are very important for individual counties within the region. *A comparison of the current LQ value of each county with whole Ohio will show the importance of these clusters in the regions based on LQ value>1.*

## Shift-Share Analysis for Ohio

The shift-share analysis helps to identify the impact of region-specific factors on regional economic growth. It explains the changes in a region’s economy (employment) by decomposing the actual changes into its three sources: a national share effect, an industry cluster share effect, and a regional share effect. For the clusters in which the region has a positive impact, the region has a competitive advantage. The shift-share analysis for the Ohio state is presented in Table 2 (appendix). Industry clusters in the table are organized by their location in cluster bubble chart. From 2015-2018, the employment for the emerging cluster Utilities at the national level decreased. However, the regional share is 6030 jobs for utilities. Ohio’s utilities cluster performed much better than the industry cluster (loss of 107,455 jobs) at the national level. The positive regional share indicates the Ohio economy enjoys a competitive advantage in Utilities, and the same case goes for wholesale trade (star cluster) and mining cluster (transforming cluster), as shown in Table 2. All in all, Ohio has a competitive advantage over wholesale, mining and utilities cluster.

## Industry Cluster Analysis for Adams County, Ohio

The transforming clusters of Ohio states don’t fall in Adams’s county. Adams County has a competitive advantage in clusters: Information, Real trade and utilities over Ohio state. The competitive advantage of Ohio in cluster “utilities” seems to be more concentrated in Adams County (Table 1, Figure 3).

Table 1. A table comparison of LQ for Adams county with whole Ohio State in 2018

|  |
| --- |
|  |

The cluster: utilities and educational services gets more attention in the cluster bubble chart (Figure 3). Utilities cluster falls in Mature cluster (LQ=6.3) in Adams County, where Adams county employment fraction is very high compared to the national level and employment has slightly decreased over 3 years.

|  |
| --- |
| A screenshot of a cell phone  Description automatically generated |

Fig 3. Industry Cluster Bubble Chart for Adams County, Ohio, US

Utilities cluster is more specialized compared to the national level. Also, the cluster: educational services is one of the essential emerging cluster in Adams county where the fractional employment is less compared to the nation but has increased rapidly (200%) over three years than the national average. It is less specialized but may become specialized in the future.

|  |
| --- |
| A close up of a map  Description automatically generated |

Fig 4. Industry Cluster Bubble Chart for Adams County, Ohio, US (two industry excluded)

Star clusters:

* manufacturing (largest in start), whole trade
* specialized, increasing concentration

Mature clusters:

* Utilities (highest LQ), retail trade (largest cluster in Adams county), healthcare
* Specialized, decreasing concentration

Emerging clusters:

* Finance, Construction (57.1%), Administrative (69.6%), Educational services (200% emp increment > national)
* Not specialized, the possibility of specialization, increasing concentration

Industry clusters are a framework for economic development that changes the focus from viewing the regional economy as many individual firms influenced by specific events to viewing the regional economy as a collection of clusters. Industry cluster emphasizes on working together as a group of firms and sharing common problems and benefits.

# DISCUSSION ON CHALLENGES

﻿Cluster analysis using location quotients and shift-share analysis provides useful information about the regional economy’s strengths and weaknesses. It can be improved by using more relevant data (primary).

Initially, I had to spend a few days on research to understand the concept of cluster analysis. Since it was the final week, time management was the first challenge. During the exam time, I utilized time on reading papers, watching YouTube videos and other sources from the internet. Though I could prepare the required visualization on Tableau spending less time, I wanted to create similar interactive visualization in Python. Therefore, I had to research on interactive visualization on Python. At first, I started working with data from a source on the internet (<http://www.statsamerica.org/innovation/anydata/>). It made my task more manageable. However, data didn’t match with the data on the Bureau of Labor Statistics. For initial analysis and coding, I copied data on excel, prepared data for cluster analysis, and then started coding for visualizing and interpreting the results. After the initial setup, I used web scraping modules in Python to scrape data from the Bureau of Labor Statistics’ website. As data were similar, I used the same excel data to prepare visualization. The date was limited to 2015 up to 2018 on the page where I was looking, because of which, I couldn’t make 5 years analysis. I just needed to showcase my skills, so I moved on from there.

The data task has helped me to think of cluster analysis in parallel to the concepts of economics and supply chain. I was able to utilize the research skills and data analysis skills that I gained from my graduate degree.

# REFERENCES

Data Browser: StatsAmerica. Statsamerica.org. http://www.statsamerica.org/innovation/anydata/. Published 2020. Accessed May 5, 2020.

QCEW Data Views. Bls.gov. https://data.bls.gov/cew/apps/data\_views/data\_views.htm#tab=Tables. Published 2018. Accessed May 6, 2020.

# APPENDIX

Table 2. Ohio Industry Cluster Shift Share Analysis

