

Colton Proctor

Homework Week Seven

Dr. Aiman Darwiche

Grand Canyon University

Homework Week 7 Colton Proctor

The data set that I chose to investigate through this assignment is one containing data pertaining to the current semi-conductor shortage. The data begins on the first of January 1985 and contains monthly index numbers all the way to November of 2021. The data for each month are ten columns, which are different indices that relate to the import and export of goods in the United States. There are also indices that include the semi-conductor industry as well as data that does not. It also contains the employment index for people working in the semi-conductor industry.

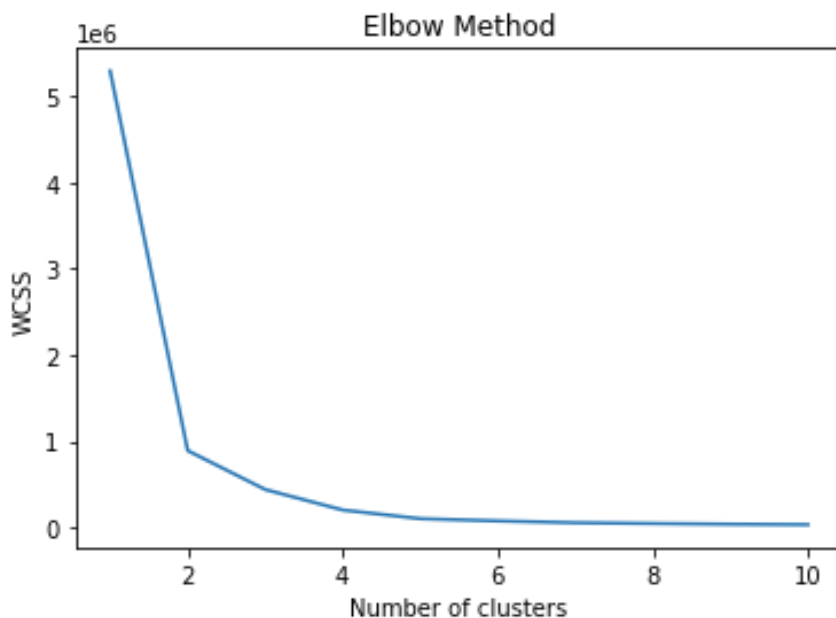
This is an important data set since there are going to continue to be shortages for semi-conductors from this point forward unless something is done. Increasingly more devices require the use of semi-conductors today than ever before, and the number is expected to rise exponentially (Markets, 2021). My goal then with this data set is to find relationships between these different indices and metrics to see how in the future the shortage can be mitigated.

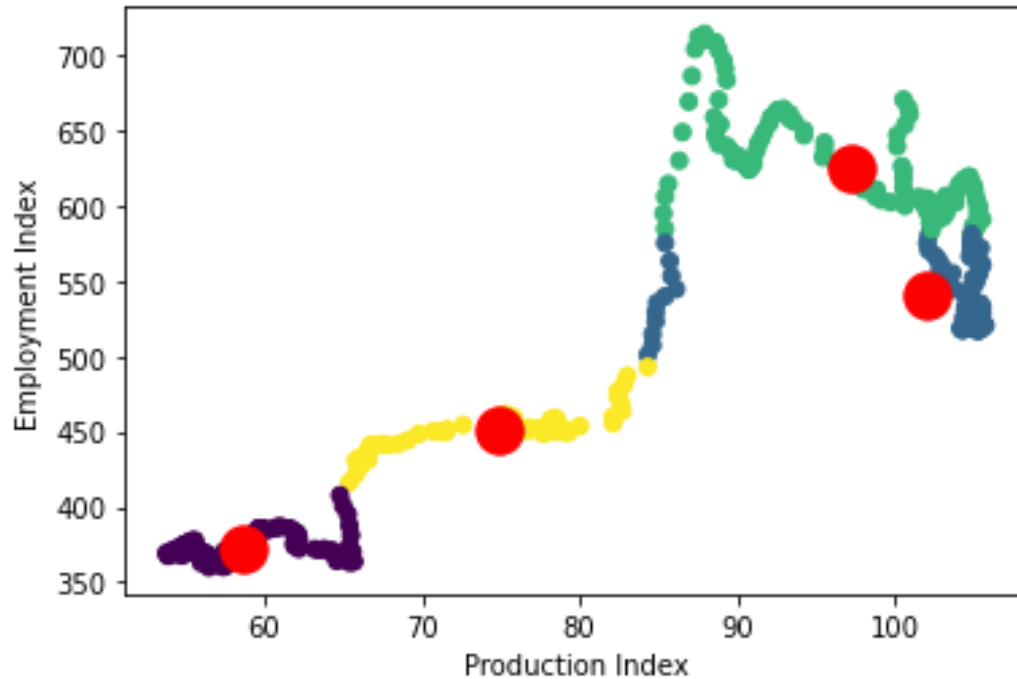
The first relationship I want to explore is that between the number of people employed in the semi-conducted industry and the producer price index for semi-conductors. This is important to show the relationship between how many jobs there are manufacturing semi-conductors to how much money is being made by exporting those goods. The more money being made from the industry would indicate that there is a surplus of semi-conductors since they are being sold instead of used in domestic production.

The second relationship is between the producer price index and the relative importance of the semi-conductor industry. That is the actual percentage of semi-conductors that are produced in relation to the total industrial production index. This is important information because it can show the quality of the semi-conductors that are being created. As well as the number of exports in relation to the number of semi-conductors produced in the country.

Third is the relationship between the relative importance of the semi-conductor industry related to the industrial production index. This will show how many semi-conductors are being purchased outside of the United States. According to an article on the times of India website due to problems within the country regarding the pandemic relying on sources such as India or China for semi-conductors is going to remain unstable for the foreseeable future.

To aid in finding the proper number of clusters to use for each of the questions I used the elbow method. For the first problem I found that using two clusters didn't give a good indication of the different relationships. Due to the data being time series data there are general trends that the model strongly gravitates toward finding. However, when the number of clusters was raised an interesting connection between the two different time periods was found.

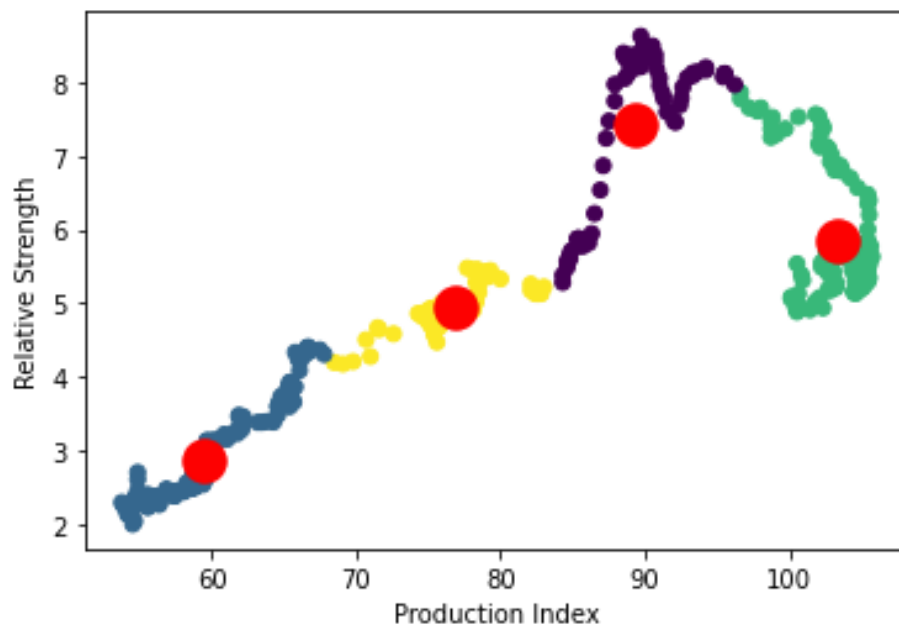
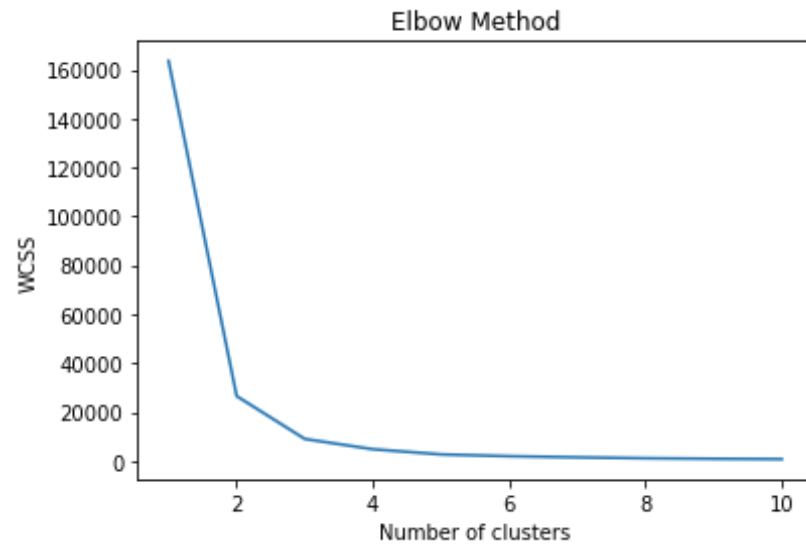




This graph shows that the number of employees is strongly related to production index for semi-conductors. This is intuitive since the more workers there are more will be produced. An interesting that I see toward the top of the graph is that there seems to be a point where having more employees didn't result in more production. I would venture that this comes with the improvement of the technology that creates the semi-conductors requires less workers to be efficient. The grouping of the high production low workforce with the blue cluster though means to me that the lower number of employees isn't necessarily the sweet spot when it comes to employment. It seems that the green cluster with higher employment has a better chance of raising production index values.

Homework Week 7 Colton Proctor

For the second question I again used the elbow method to determine the proper number of clusters.

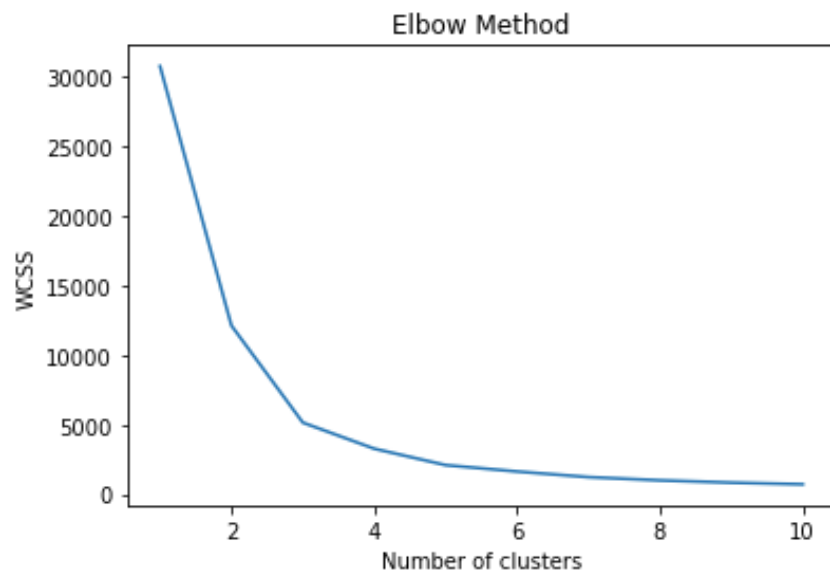


This graph shows the relationship between the relative strength of the semi-conductor industry against the production index, which includes the amount of manufacturing including semi-conductors that is being done in the United States. It seems that when the market there are a lot of semi-conductors being used, that the amounts that are being produced in the United States

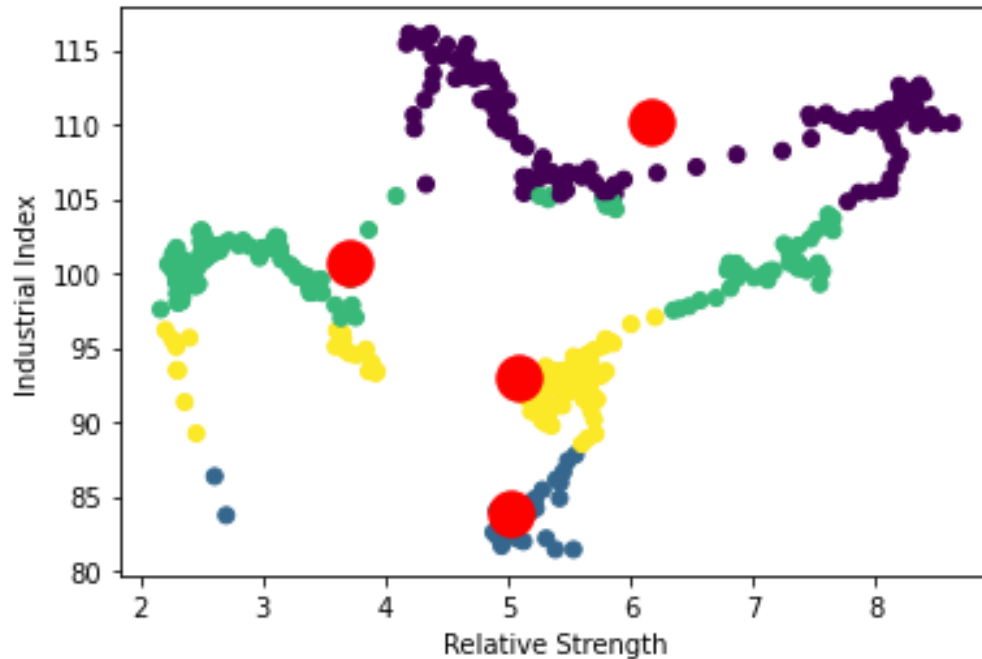
Homework Week 7 Colton Proctor

also rises. There is a high concentration at the very top end of the production index where the market share stagnated around 5%. This means that there is also a large amount of importing that is happening. Overall, it seems that increasing the amount of production however increases the overall production index as well. As seen in the first question this means that more workers would be a good answer to both problems.

For the third question the elbow method was again utilized.



This data was different in that the elbow was slightly right skewed related to the other two graphs. It seemed that the sweet spot was either three or four clusters. I ended up using four clusters because I felt it captured the chaotic relationship a little more closely.



The spread between the relative strength can be explained by the number of imported semi-conductors used to facilitate the industrial growth. However, I find it interesting that the clustering algorithm ended up grouping together different points more closely related to their industrial index over the relative strength. There is a high concentration however where the high relative strength related to a high industrial index. It seems that, the more workers that are employed, the more chips are produced locally, the more manufacturing is possible overall.

The ethical aspects of using open-source data depend on the context that they are used. The data set that I chose is using publicly available index values which are posted monthly to try and find trends to prevent semi-conductor shortages in the future. To me that is a good application of publicly available data, as it is for the greater good and openly accessible. If the data were to say be leaked from internal documents after a company was hacked, that's a different story. I think however that forums like Kaggle are a step in the right direction for humanity. It's a place where people can aggregate data for others to use and try to find answers.

Works Cited

Markets, R. and. (2021, December 15). Global Semiconductor Chips Shortage Analysis Report

2021. Retrieved December 16, 2021, from <https://www.prnewswire.com/news-releases/global-semiconductor-chips-shortage-analysis-report-2021-301445467.html>

TIMESOFINDIA.COM / Updated: Dec 15, 2021. (n.d.). *Semiconductor shortage: Cabinet*

approves rs 76,000 crore incentive plan to woo manufacturers - times of India. The Times of India. Retrieved December 16, 2021, from

<https://timesofindia.indiatimes.com/business/india-business/semiconductor-shortage-cabinet-approves-rs-76000-crore-incentive-plan-to-woo-manufacturers/articleshow/88298699.cms>