Alisha Khan

TE360 Final Project Report

2nd December 2020

OVERVIEW

This project will analyze the correlation between the increase in incubators/accelerators and employment levels in the Chicago metro area.

GOALS

1. Analyze datasets from the Illinois Department of Employment Security and determine the relationship between employment rates and the creation of incubators/accelerators
2. Observe how employment levels change different sectors
3. Create a simulation for how overall employment rates, as well as employment rates per sector, will change depending on the number of incubators/accelerators

HYPOTHESIS

As the number of incubators/accelerators increase in Illinois, so will the employment levels in the Chicago metro area.

I came up with this hypothesis while doing one of the readings for TE360. The following articles were very interesting to me: Forbes Article, Brookings Article. As I read through them, I was wondering if it was possible for me to find the correlation between employment and the increase in incubators/accelerators since it seems that it has a positive impact on the economy.

Incubators and accelerators help out with the beginning of building a company, therefore it is plausible to believe that they would also increase employment since they play a role in the creation of more jobs.

It is hard for there not to be confounding variables, but I still believe that the increase in incubators/accelerators would have an impact on employment.

MODEL

I used the following sources as input.

1. The Illinois Department of Employment Security has released reports every year since 2001 titled *Where Workers Work*. This report features private sector employment totals for six northeastern Illinois counties. It also highlights the overall employment totals for these counties, and how they have changed since the prior year.
    a. Links: 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
2. Intersect Illinois produces a comprehensive report of Illinois' ecosystem. I looked through this report and counted the number of accelerators and incubators. I then searched up when each of them was founded and made a table of the cumulative total of incubators/accelerators from 2001 to 2019.

I had to do a lot of data cleaning with the Illinois Department of Employment Security Reports. It was a little challenging since I am not very familiar with Pandas or Excel. Since it was a pdf file, I had to convert this to a CSV. After doing so, I was able to do data cleaning in both Python and Excel.

Then, I added a column to the CSV for the cumulative total for incubators and accelerators. I noticed that 2015-2017 were the only consecutive years that each had increases in the number of incubators. I decided to base my model on only those three years so the equation would not have a very shallow slope.

This is what one of my dataframes looked like after the cleaning process. I was not able to include the whole data frame as there were many columns with the different industries.

Out[7]:

	YEAR	INCUBATORS	ALL INDUSTRIES	AGRICULTURE, FORESTRY, FISHING, & HUNTING (11)	MINING, QUARRYING, & OIL AND GAS EXTRACTION (21)	UTILITIES (22)	CONSTRUCTION (23)	MANUFACTURING (31-33)	FOOD	BEVERAGE & TOBACCO	...
0	12/31/15 0:00	16	3444928	2345	1115	10616	123323	357915	45247	4004	...
1	12/31/16 0:00	17	3507834	2768	1205	10750	132125	358032	46752	4221	...
2	12/31/17 0:00	19	3555979	2779	1166	11139	133567	358190	47763	4960	...

3 rows × 44 columns

Then, I started the modeling and fitting process. I imported the following libraries: Pandas, NumPy, Matplotlib, and Seaborn. Pandas allowed me to convert the CSV files into dataframes. Using the dataframes, I was able to do a lot with the data.

I used the polyfit method from NumPy to determine a least-squares polynomial fit. I felt that this was more appropriate than using a simple linear regression because I was not sure about the different patterns in my data. I knew that overall in the Chicago metro area there was an increase in employment, but I did not know the specifics for each industry. That is why I felt this fit would be more accurate for my data.

Using the polyfit method, I was able to plot this equation against the actual data points and display this using seaborn.

Using this method I was also able to display estimates of total employment when there is an *X* number of incubators/accelerators.

I think the weakness of this model is that there are many confounding variables. There is no way for me to say that the employment increase is the direct product of the increase in accelerators/incubators. Also, not every incubator and accelerator is the same "size". Some are smaller and take on fewer companies than others, so that makes this model less accurate.

Another weakness is that I was not able to separate incubators and accelerators, this is because of the lack of data on them. That is why I chose to just make it a group and have the group be the sum of incubators and accelerators.

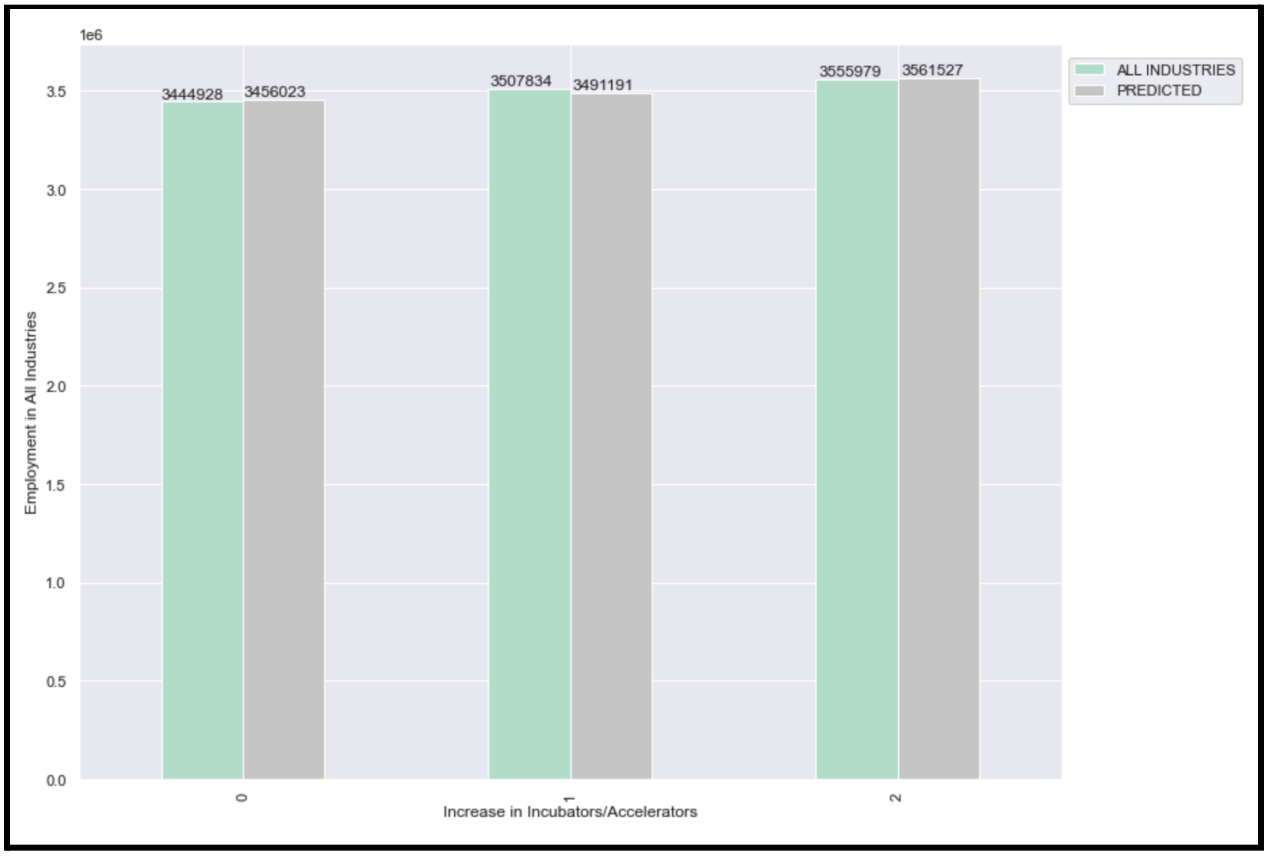
OUTCOME

The data matches with my hypothesis that as the number of incubators/accelerators increase, so will the employment levels in the Chicago metro area.

I think the model I made is very useful as well. There are two things it can do, produce a graph, and produce an estimation. It can make a graph of any county and industry. It can also predict employment in a specific county and industry when there is a certain number of accelerators and incubators.

The graph makes it easy to see how employment changes. In some combinations of counties and industries, I even saw a decrease. For example, Will County and the Computer and Electronic Product Manufacturing subsector.

The following is the output from my model. The least-squares polynomial fit seems to have been a good choice as the predicted value is very close to the actual value.



This outcome shows that it is very likely for employment to increase if the number of accelerators and incubators increase.

I think that we have a lot of great accelerators and incubators in Chicago and that this data shows that they most likely have a role in helping out the economy by creating more jobs. I think if the accelerators and incubators continue to form at the pace they currently have been, employment will increase steadily.

With COVID going on, and lots of people losing their jobs, I think that we do need to take a more aggressive approach with reviving the economy and getting people jobs, so if there was an initiative to start more incubators and accelerators, I think that would greatly increase employment.