DATA COLLECTION AND CLEANING

While collecting and cleaning the data, a few wacky things prevented us from having a straightforward process. The first being that the website changed the formatting of how they identified industries a few years into the process. They changed from SIC standards to NAICS standards, so using data from before 1998 wasn't a possibility. Another problem that occurred while collecting data was that every 5 years a new standard of NAICS standards came out, and the people that collect the CBP data decided it was best to change the column name, instead of leaving it a general "NAICS CODE" tag. Lastly, this all had to be formatting into different URLs to request instead of receiving all the data in a nice JSON file and parsing it that way.

RESEARCH QUESTIONS

After messing with the data, seeing what we had, and seeing what we didn't have. We decided to consult our original proposal for what questions we could bring and likely predict/attempt to predict. Our data best described industries and how they changed over time, including their employment numbers, how many companies there are in an industry, and what the annual payroll looks like for that industry. This available data lead us to choose research questions like: How did the job landscape change throughout the years (Did it become more white collared or blue collared)? Which industries will likely see a monopoly emerge in the future or at the moment? And what industries will likely see the fall of a monopoly (Will barriers to entry change)?

MACHINE LEARNING APPLICATIONS

Out of the 3 counties analyzed, we used linear regression to predict whether or not some 3 of the most successful industries in terms of revenue generated in 2018 would monopolize or yield in the future.
We will be comparing the linear regression graphs of the number of establishments in a county vs the number of employees in that industry over the time frame of 1998 all the way till 2018.

San Mateo County Observations:

1. There is a boom in the information services in San Mateo County which could be explained due to its close proximity to the Silicon Valley. The IT industry clearly shows no signs of being monopolized.
2. There is a rise in the scientific services industry, showing a rise in the number of white collar jobs, but a dip in the number of employees which could signify the need for specialization in this industry..
3. There is a decline in the wholesale industry but a rise in the number of employees which clearly shows how companies like Walmart and Costco have

monopolized this industry. The linear regression model predicts the wholesale market to be monopolized.

Los Angeles County Observations:

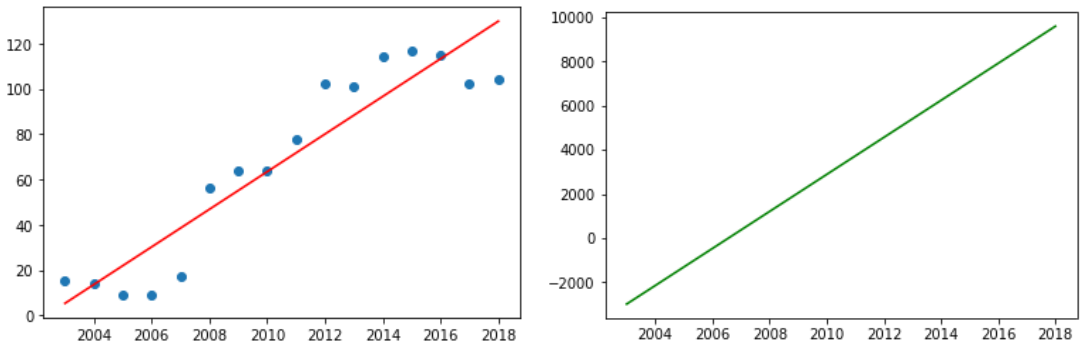
1. There is a rise in the scientific services industry, showing a rise in the number of white collar jobs, but a dip in the number of employees which could signify the need for specialization in this industry.
2. There is a dip in both the number of wholesale companies and the number of people employed in this industry which shows that there has been a decline in this industry, a sign showing that blue collar jobs are becoming less available in LA county.

San Luis Obispo Country Observations:

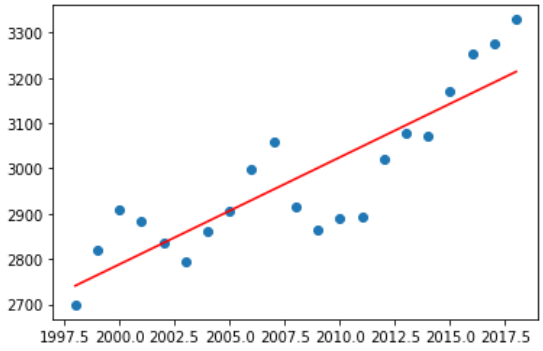
1. There is a rise in the number of establishments and number of employees in the Food and Services industry which clearly shows a growth in the number blue collar jobs in the county as this industry generates a significant revenue in the county.
2. The utilities industry has also seen a rise both in the number of establishments and people employed, signifying that there are more blue collar jobs still available in the county.

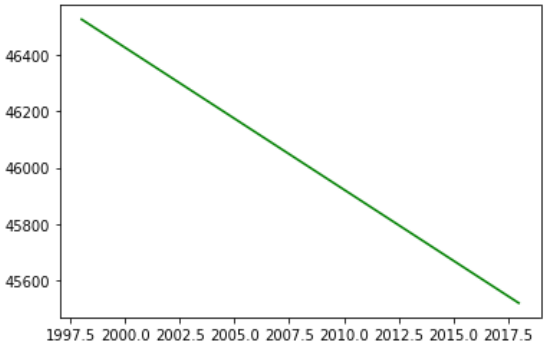
SAN MATEO COUNTY

Information Services

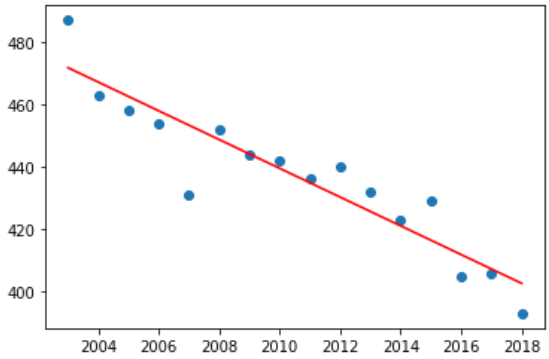


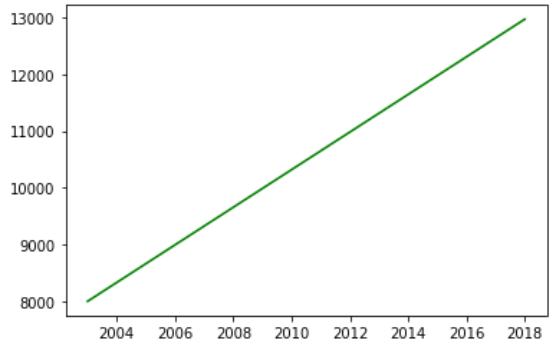
Professional, scientific, and technical services





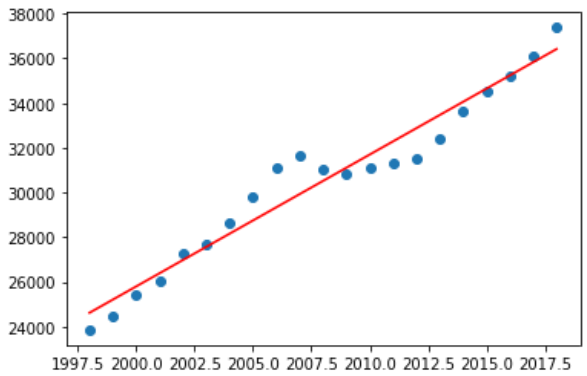
Wholesalers

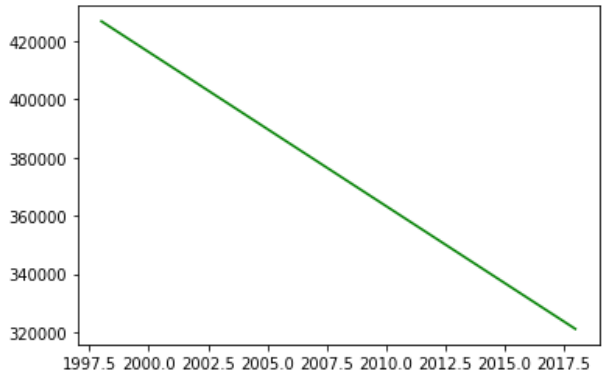




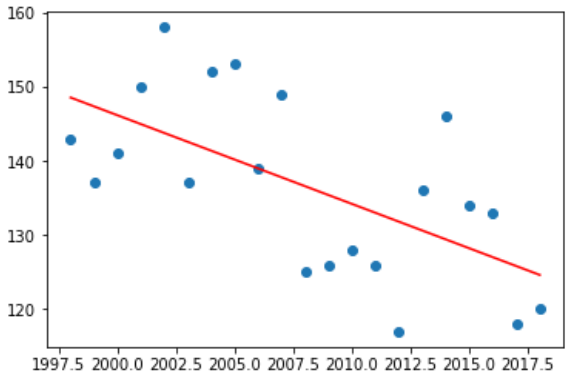
LOS ANGELES COUNTY

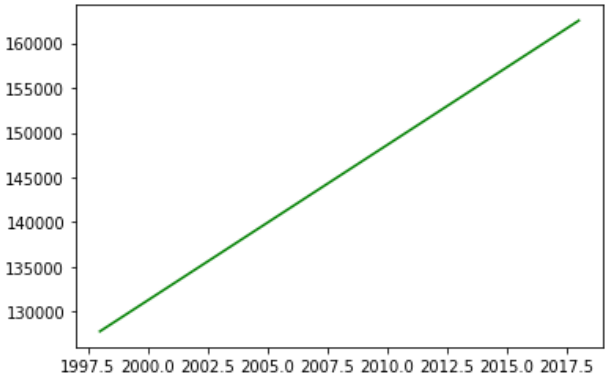
Profession, scientific, and technical services



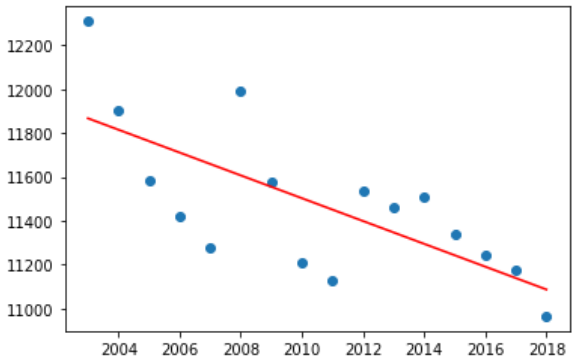


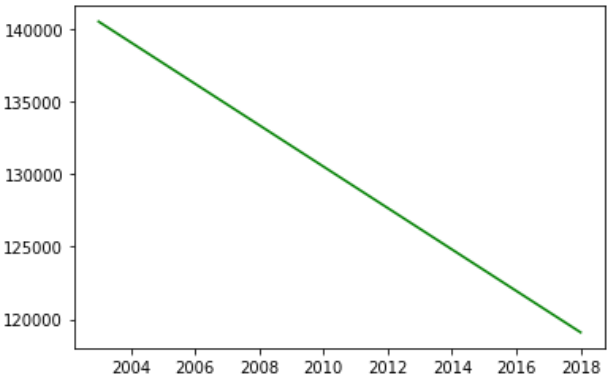
Hospitals





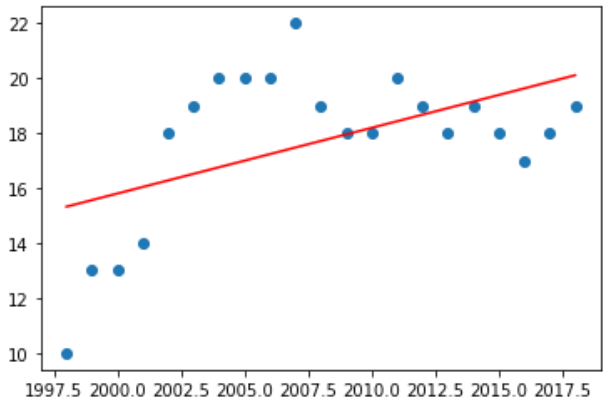
Wholesalers

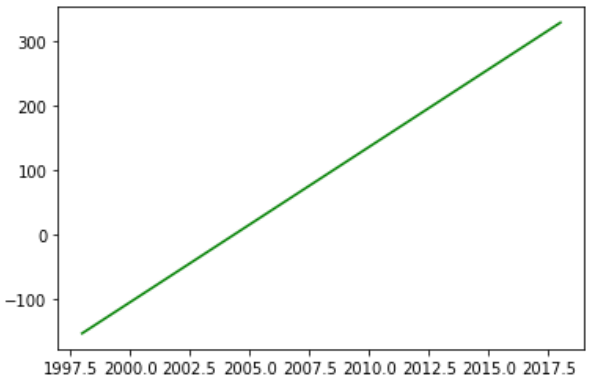




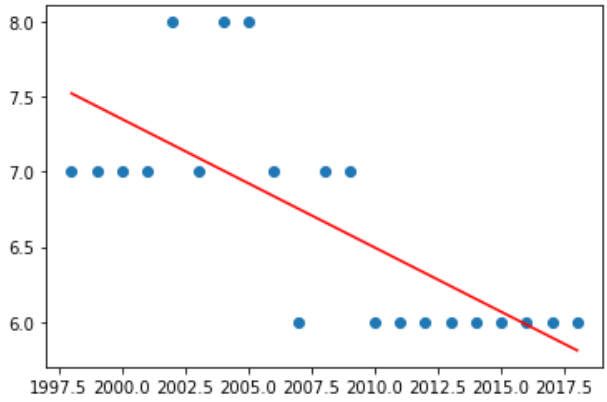
San Luis Obispo County

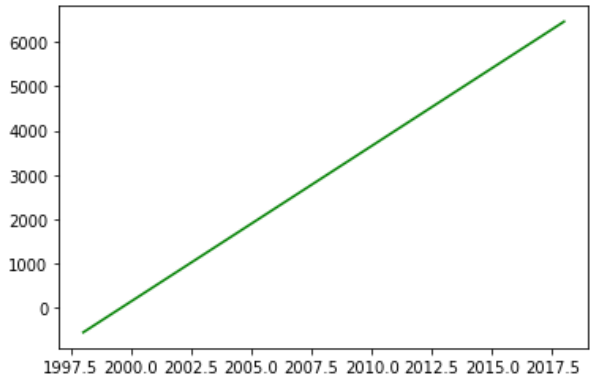
Utilities





Hospitals





Food Services