Capstone Project on CPI and Employment

# Team

Debabrata Kabiraj  
Joseph Simone  
Rajwant Mishra

Github: <https://github.com/josephsimone/DATA698>

# Motivation

We are in totally different era of twenty first century, and it gives us very rare situation where any positive news would help the humankind. We want to use the historical CPI data and find the relation of it with the employment, in hope that we would have some positive news on employment by following the trend of the data in past.

We feel that lower CPI would result in more job opportunity, as the it gives space for more competition in small business across sectors.

# Research Method

## Types of Research Data

We will use existing data available from US government to do this analysis, with some aggression of the info to derive the missing info.

## Research philosophy

We don’t plan to change or replace any of the data points to meet our goal. Since data is used as is with some grouping based on the predefined key, value pair we expect we would not add any bias in this process. We would use the data to amylase the fact about the data.

## Types of Research

Since, the analysis knowledge gained would not add immediate practical implications of the finding, we feel it’s Fundamental research on the topic to bring the correlation and biasness of these two disjoined data points.

## Research approach

We would be following an *inductive* approach, as we would like to find answer to specific research question(s) formulated in the beginning of the research process. Additionally, we will be also follow *deductive* approach, since we have chosen to achieve research objective(s) via testing hypotheses.

## Research design

We would be doing following *Conclusive* research design, which provide final and conclusive answers to the research question.

## Sampling

We would divide data in Train and Test set. Test set would include data from 2017 to till date data available. Train set would be further sampled on 70:30 (Train : Test) ratio in multiple Random sampling process.

## Missing value

We would treat missing value as the average of last years and following years value only, instead of taking the whole years average. This would keep impact of missing value limited to neighboring years only.

## Ethical Considerations

We would not infer any info apart from the given set of data point. We would try to properly take care of the outlier’s impact on the data and report it. We will avoid any type of misleading information, as well as representation of primary data findings in a biased way.

# Data Definition

**CPI:** The consumer price index (CPI) measures the average level of prices of goods and services in the economy. The CPI formula is used to measure the change in prices by consumers for a representative basket of goods and services during a defined time period. CPI is a widely followed measure of inflation which is used by economists, policy makers, investors to guide economic policy, forecasting and investment decisions.

1. *What is the CPI Formula?*

CPI = Cost of market basket in a given year / Cost of market basket in base year ) x 100

1. *How is the CPI market basket determined?*

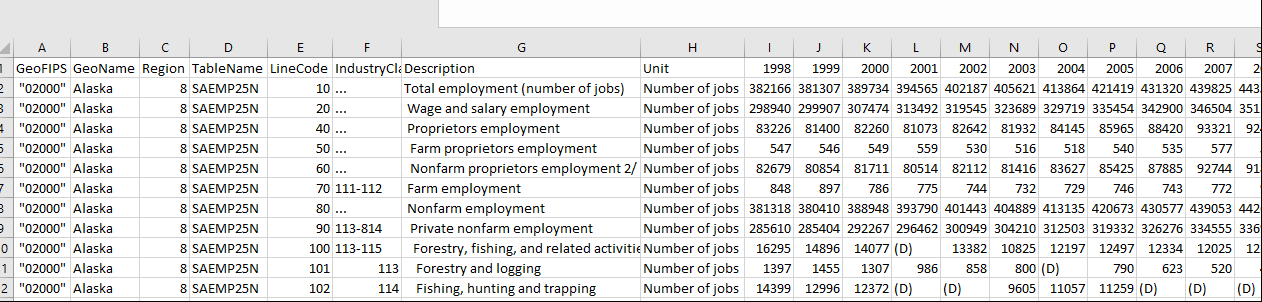
The CPI market basket is developed from detailed expenditure information provided by families and individuals on what they actually bought. There is a time lag between the expenditure survey and its use in the CPI. For example, CPI data in 2016 and 2017 was based on data collected from the Consumer Expenditure Surveys for 2013 and 2014. In each of those years, about 24,000consumers from around the country provided information each quarter on their spending habits in the interview survey. To collect information on frequently purchased items, such as food and personal care products, another 12,000consumers in each of these years kept diaries listing everything they bought during a 2-week period.

Over the 2 year period, then, expenditure information came from approximately24,000 weekly diaries and 48,000 quarterly interviews used to determine the importance, or weight, of the item categories in the CPI index structure.

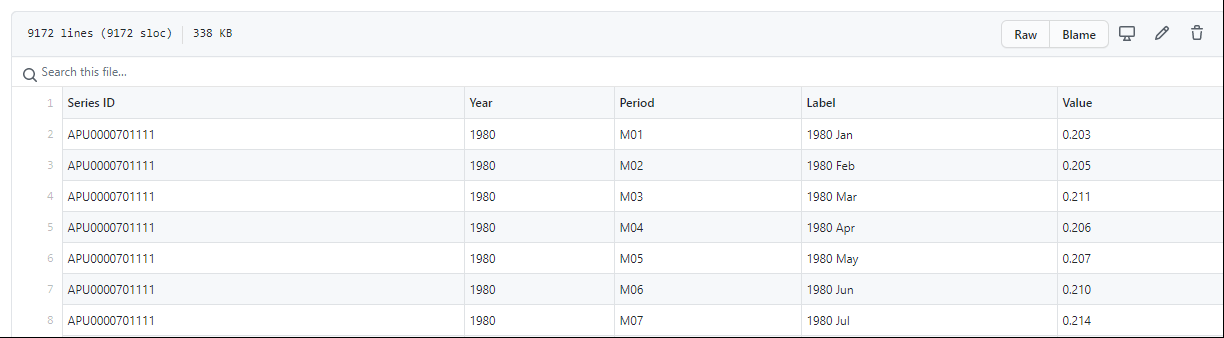
**Employment By Industry**: The number of workers earning a wage or salary, along with those who are self-employed. These statistics count workers in private industries and in federal, state, and local government.

Data tables produced annually for employment by industry include: the number of full-time and part-time employees in each industry, full-time equivalent employees, self-employed persons, and persons engaged in production (full-time equivalents plus the self-employed).

# Sample Data:

Sample of Employment Data:

Sample of Consumer Price Index (CPI):



# Hypothesis

We feel CPI has some impact on the Employment Data by each sector and national level. We would validate this with our analysis in the project.

**Null Hypothesis**: Consumer price index (CPI) has no relation with Job market (working Employment Data).

**Alternate**: Employment is related to CPI.

**Other Question that we wish to answer by our analysis:**

1. Which CPI parameter has most impact on the Job market ?
2. What are TOP 5 sector that gets most impacted by CPI ?
3. Which sector has no impact of CPI?

# Appendix:

https://www.bls.gov/cpi/data.htm

<https://www.bls.gov/cpi/>

Extra Ref: <https://www.sciencedirect.com/science/article/pii/S1877050912002591>

<https://www.realogicinc.com/understanding-cpi-in-commercial-real-estate/>

<https://www.nber.org/papers/w4505> (The Consumer Price Index as a Measure of Inflation)

<https://research-methodology.net/research-methods/>

<https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/datopic.html#:~:text=Data%20Analysis%20is%20the%20process,and%20recap%2C%20and%20evaluate%20data.&text=Indeed%2C%20researchers%20generally%20analyze%20for,Savenye%2C%20Robinson%2C%202004>).