

Paul Brennan

Professor John Coles

CIS 512 – Intro. To Data Science and Analytics

November 6<sup>th</sup>, 2016

## Initial Analyst Workflow Plan – Vermont Economic Stimulus

### **Core Question**

Taking a broad stance, we are attempting to formulate a plan to increase economic activity in Vermont with an inflow of \$10 million. However, this stance needs a lot of clarification, namely in the desired results. Although there are competing methods to measuring economic activity and quality of life, GDP seems to be the defacto standard, with the advantage of offering objectivity and consistency in calculation. (“How Do We Measure ‘Standard of Living’”) If we are going to be measuring success terms of increase to GDP, it would be apt to understand what GDP really represents, and how it is measured. According to the Bureau of Economic Analysis, GDP is a total measure of production, be it goods or services, created for consumption. (McCulla, 2-4). Goods and services used as inputs for further production are excluded from the calculation to prevent double counting of value. This caveat is for excluding things like raw materials and their associated labor, rather than things like tools and machines used in the production process. The point is to capture total value generated by an economy through it’s end products. This also means that items produced, but not yet consumed, count towards GDP.

The easiest way to conceptualize results is by comparison. With a pool of \$10M, it would seem the least we could do is add \$10M to the state’s GDP for one year. The worst we could do is spend \$10M, and create no appreciable increase in GDP. More ideal scenarios would entail a sustained increase to the state’s GDP, implying a lasting impact on economic activity.

Currently, our approach focuses on the information technology sector, as that is a field often brought up as a high growth area, yet initial research suggests it's share of GDP has been shrinking in Vermont from 2006 to 2014. Our initial strategy is to foster co-operation between the shrinking IT sector and more resilient industries. We will need to identify strong industries to target. We will need to identify what regions of Vermont we should focus our efforts on. We should identify what the existing workforce pool looks like, and if we can leverage it for job creation, or if we will need to import the workforce. Likewise, we should do the same for businesses and other sources of capital. We will should try to identify the potential problems the the IT sector has faced in the past.

## **Data Ingest**

Data on the GDP breakdown by industry for state (VT), region (New England), and the nation as a whole should be obtained. Worker income, business income, and employment breakdown may also prove to be useful. Population statistics should be obtained, broken down into regions (county, municipality, census tract), which would provide information on the workforce, such as education and employment.

## **Data Preparation**

Many of these data sources are provided as tables with sub-categories in the columns and years in the rows. We will likely have to carefully scrape them to turn them into a more useable format.

## **Basic Data Analysis**

To show our approach is a sound concept, several things must be confirmed: How does growth in Vermont's IT sector compare to the region, and the country as a whole? How does this compare to other industries? Do these trends vary significantly based on metric? (i.e. How closely are GDP, employment rate, and employment compensation correlated?) These questions should be explored on

from the state as a whole, and from regions within the state. We should also try to correlate large shifts with real world events, such as a factory opening or closing, or the 2008 recession.

Due to the nature of this data, we will likely not need to rely on more advanced methods. Most methods will be time-series analysis, looking at trends of different metrics over time. Other analysis will be accomplished by simple numerical ranking. This work will likely be done in some combination of python and R, using python for data cleanup, and R for ease of calculation and graphing.

## **Preparation**

The simple analysis will be reflected in simple preparation. Trends will be represented with simple plots containing a best fit line. Some data may be represented as tables. The most complicated display would likely be a map, showing regional breakdown.

## **Reproducibility**

The key to reproducibility is good documentation. Strong use of scripting for manual tasks should aid this as well. If done right, the methods used should be repeatable by another party, and should even be easily adaptable to other states or regions. Much of the data that is planned to be sourced is from national agencies, which should lend well to adaptability.

## **Presentation**

Presentation will be highly dependent on the results found. However, the in person presentation and the report should cover the following areas:

1. Description of the problem
2. Analysis of Trends, including associations
3. Industry Profiles, of IT sector and candidate related sectors
4. Proposed action plans

## Bibliography

"How Do We Measure Standard of Living?" (n.d.): n. pag. Federal Reserve Bank of Boston,

13 Feb. 2003. Web. 5 Nov. 2016.

<<https://www.bostonfed.org/education/ledger/ledger03/winter/measure.pdf>>.

McCulla, Stephanie H., Shelly Smith, and Brent R. Moulton. "Measuring the Economy: A Primer

on GDP and the NIPAs." (n.d.): n. pag. Bureau Of Economic Analysis, 15 Mar. 2016. Web.

5 Nov. 2016. <[http://www.bea.gov/national/pdf/nipa\\_primer.pdf](http://www.bea.gov/national/pdf/nipa_primer.pdf)>.