# **Project Proposal: Group SuRe Thing!**

Group members: Rebecca Ofori Yeboah and Suzanne Thepaut-Hasselback

## **Executive Summary**

This project will use various analyses software, including Python, R and Tableau, to investigate the evolution of unemployment in the U.S. between 1980 and 2020. Our focus will be significant spikes in unemployment over this time span, as compared to the peak in 2020, and the variation in the unemployment rates of different sectors of the eligible working population in America. The team will follow a 6-week scrum schedule and be supported by the Entity mentoring team.

## 

## **Business Objectives**

Our objective with this project is to investigate two different angles of the evolution of unemployment percentages in the U.S. labor market between 1980 and 2020:

* Since 1980, in comparison to the COVID-influenced spike in 2020, what other sharp increases are visible in overall annual unemployment percentages? What events do these peaks correspond to? Are these peaks’ percentage values significantly different from the value of the peak in 2020?
* Do the unemployment peaks vary when looking at different sectors of the population? Were some sectors’ values very different from the average unemployment values of the time? Are these variances consistent with other peaks in the graphs over time? (Did high unemployment events hit every sector equally hard?)

## **Background**

Since 2017, a major global slowdown in growth has been taking place in many countries of the world, and several European governments have had economic crises. These measures began to be exacerbated in September 2019, when the Federal Reserve had to begin interventions in the repo market after the overnight lending rate spiked above the Fed's target rate in an attempt to keep the economy afloat due to a liquidity issue. The Fed's efforts began to fail when the first documented case of COVID-19 emerged in Wuhan, China in November 2019. The government in China first instituted travel restrictions, quarantines and stay-at-home orders. When efforts to contain the virus in China were unsuccessful, other countries instituted similar measures in an attempt to contain and slow the spread of the virus, which prompted many cities to close. The initial outbreak expanded into a global pandemic. The economic effects of the pandemic were severe. More than 24 million people lost jobs in the United States in just three weeks. Official economic impact of the virus is still being determined but the stock market responded negatively to the shock to supply chains, primarily in technology industries.

This has raised the question of whether the nation will fully recover from this global recession. A look at the historical data might provide some insight, given that there have been about four (4) recessions since 1980. Some did not impact as many people as others, whereas others impacted a particular demographic more.

In the year 1980, the National Bureau of Economic Research (NBER) considers a very short recession to have occurred, followed by a short period of growth and then a deep recession. Unemployment remained relatively elevated in between recessions. The recession began as the Federal Reserve, under Paul Volcker, raised interest rates dramatically to fight the inflation of the 1970s. The early 1980s are sometimes referred to as a "double-dip" or "W-shaped" recession.

Then in 1981-82, the Iranian Revolution sharply increased the price of oil around the world in 1979, causing the 1979 energy crisis. This was caused by the new regime in power in Iran, which exported oil at inconsistent intervals and at a lower volume, forcing prices up. Tight monetary policy in the United States to control inflation led to another recession. The changes were made largely because of inflation carried over from the previous decade because of the 1973 oil crisis and the 1979 energy crisis.

After the lengthy peacetime expansion of the 1980s, inflation began to increase and the Federal Reserve responded by raising interest rates from 1986 to 1989. This weakened but did not stop growth, but some combination of the subsequent 1990 oil price shock, the debt accumulation of the 1980s, and growing consumer pessimism combined with the weakened economy to produce a brief recession.

The 1990s were the longest period of economic growth in American history up to that point. The collapse of the speculative dot-com bubble, a fall in business outlays and investments, and the September 11th attacks, brought the decade of growth to an end. Despite these major shocks, the recession was brief and shallow.

The subprime mortgage crisis led to the collapse of the United States housing bubble. Falling housing-related assets contributed to a global financial crisis, even as oil and food prices soared. The crisis led to the failure or collapse of many of the United States' largest financial institutions: Bear Stearns, Fannie Mae, Freddie Mac, Lehman Brothers, and AIG, as well as a crisis in the automobile industry. The government responded with an unprecedented $700 billion bank bailout and $787 billion fiscal stimulus package. The National Bureau of Economic Research declared the end of this recession over a year after the end date. The Dow Jones Industrial Average finally reached its lowest point on March 9, 2009

In light of the economic devastation, researchers found that some demographic groups and industries suffered more than others. For instance, since the U.S. Bureau of Labor Statistics started collecting data on the African American unemployment rate in January 1972, this rate has more often than not been twice as high as the white unemployment rate. In fact, between January 1972 and December 2019, other than during the aftermath of recessions, the African American unemployment rate has stayed at or above twice the white rate. The only time that the African American unemployment rate was significantly less than twice the white unemployment rate was during the Great Recession. The rate dropped after the recession’s start and lasted a few months after the technical end as the white rate increased. But even when the African American rate fell below double the white rate, it never fell very far, as African Americans experienced greater amounts of layoffs. Between January 1972 and December 2019, it never reached as low as 1 1/2 times the white rate.

## **Scope**

The scope of this project is to analyze annually reported national unemployment data from the website of the U.S. Bureau of Labor Statistics from 1980-2020. We plan to use Python, R, and Tableau, and other tools taught in our Data Science course, to complete these analyses. Our objective is to fine-tune our focus on our independent variables in order to discover significant findings. We are aiming for a visual representation of the annual unemployment numbers over time, and a statistical significance test in comparison with the general U.S. population. No other analyses are planned currently.

## **Functional Requirements**

We plan to use Python and R, Tableau, and Excel if necessary. Our dataset is sourced from the U.S. Bureau of Labor Statistics, at <https://www.bls.gov/cps/tables.htm>. Our dependent variable will consistently be unemployment percentage values from 1980-2020, and we have multiple independent variables in our study:

* Time: 1980-2020; Categorical: # of levels = 41
* National Crises/Pandemic/Event; Categorical: # of levels = 3 or 4?
* Gender; Categorical : # of levels = 2
* Race/Ethnicity; Categorical : # of levels = White/Asian/Black/Hispanic = 4
* Age; Categorical : # of levels = 7
* Occupational sector; Categorical : # of levels = 5

## **Personnel Requirements**

Personnel involved in this project are students, Suzanne Thepaut-Hasselback and Rebecca Ofori Yeboah, with help from Professor Joseph Raetano and Mr Devin Moya.

Each student allocates at least 20 hours per week and will use daily stand up forms to communicate team retrospective and sprint for the week. Each student is expected to complete a sprint and move cards on the Kanban board from the Sprint to the Done column by the end of the week on Sundays.

The team will also meet with the instructor on Mondays at 7:30 pm in addition to daily standup meetings, scheduled based on Rebecca and Suzanne’s availability. Each will complete a daily list in the shared GoogleDoc to indicate what we did the day before, what we are going to do that day, and any obstacles preventing us from meeting personal goals for that day. On Sundays at 7:30pm, the team will meet to close out the week and prepare the scrum for the coming week/hand off scrum master responsibilities. Scrum master responsibilities will trade back and forth between team members every week.

## **Delivery Schedule**

Week 1: Project planning (Scrum master Rebecca)

Connect with each other, establish communication and norms. Decide the topic, choose a dataset, define questions, and determine IVs and DVs. Research the recent history of U.S. unemployment. Set up Trello and GitHub.

Week 2: Data Wrangling (Scrum master Suzanne)

Run preliminary visual analyses to determine which 3 IVs will be retained in the second analysis. Isolate and clean up data to be used. Determine whether we have enough data to run in Python or if we need to use R. Determine if we want to control for other factors (ANOVA vs ANCOVA).

Week 3: Data Exploration (Scrum master Rebecca)

Run definitive visual analyses for comparison. Run ANCOVA or ANOVA testing, and test for significance of difference between values. Deal with any issues this causes to arise: not enough data, mis-assignment of factors, multicollinearity, etc.

Week 4: Data Analysis (Scrum master Suzanne)

Review and validate findings from the previous week, and draw insights and conclusions. Check back with the instructor to be sure we haven’t missed anything!

Week 5: Data Visualization (Scrum master Rebecca)

Create a visual presentation using analyses and findings. Verify that we haven’t missed anything. Be certain the presentation is the best it can be!

Week 6: Data Reporting (Scrum master Suzanne)

Run the presentation past family, friends, instructor, mentors… Practice presenting as a team on Zoom.

Week 7: Present to everyone!

## **Assumptions**

We are assuming that Python, R, Tableau, and other programs (including web browsers) will all be functional and not cause issues, either inherently or due to updating issues. We are assuming that our two computers will be up to the task of handling these planned analyses. We are assuming also that our analyses will not require us to use paid versions of any of the programs we will need. Finally, we are assuming that the U.S. Bureau of Labor Statistics has collected this data consistently and without errors.

## **Limitations**

Our limitations include our learning curve for GitHub and other tech tools; the possibility that the data we’ve planned on using might be insufficient for the analyses we want to do; and the possibility that weeks may overrun the planned scrum timing - not all work may get done on time.

## **Risks**

We have a number of risks in this project: there may be temporary loss of internet due to bad weather (this has already happened in Week 1); one of our computers might break for some reason; our other jobs may take more of our time than foreseen (Suzanne is a K-12 teacher and Rebecca is responsible for group home residents); and/or family emergencies may arise.

## **References**

<https://www.americanprogress.org/issues/economy/reports/2020/02/24/480743/persistence-black-white-unemployment-gap/>