**1. Introduction**

With an unsteady economic market, an increase in unemployment, and extreme measures affecting the average working-class members in the United States right now, our group wanted to investigate the intersection between economics via the S&P 500, and the unemployment rates in years 2005 and beyond (to present day), to see if there is a measurable relationship between stock market performance and the job market. By analyzing historical data on the S&P 500 index alongside national unemployment rates, our group aims to determine whether shifts in the stock market tend to precede, follow, or move concurrently with changes in unemployment. This intersection is particularly relevant today, as economic volatility directly impacts both corporate stability and individual livelihoods.

In this project, our goal is to assess whether trends in the S&P 500 can serve as an early indicator of labor market stress (unemployment), or vice versa, and to explore how these insights could inform both policy decisions and personal financial planning.

**2.Data**

To pursue this internal inquiry, we gathered data from two key sources: S&P 500 Trends from Kaggle and the U.S. Bureau of Labor Statistics (BLS). For unemployment trends.

*2.1 Unemployment Rates*

We utilized an interactive graph published on the BLS website titled “Civilian Unemployment Rate.” Unlike past class projects where we scraped entire websites, this time we focused specifically on scraping the HTML content related to the graph alone. This allowed us to extract the necessary data without the extra overhead from dropdowns, embedded links, and unrelated content. Since the BLS compiles data from federal records and national census reports, our dataset was complete and contained no missing or null (NaN) values. To prepare it for analysis, we reformatted data types and created a new column, Unemployment Change, which calculated the month-over-month difference in unemployment rates. This data was put into a data frame labeled ‘unemployment\_df’ The final structure of this dataset is detailed in Table 1: Data Dictionary.

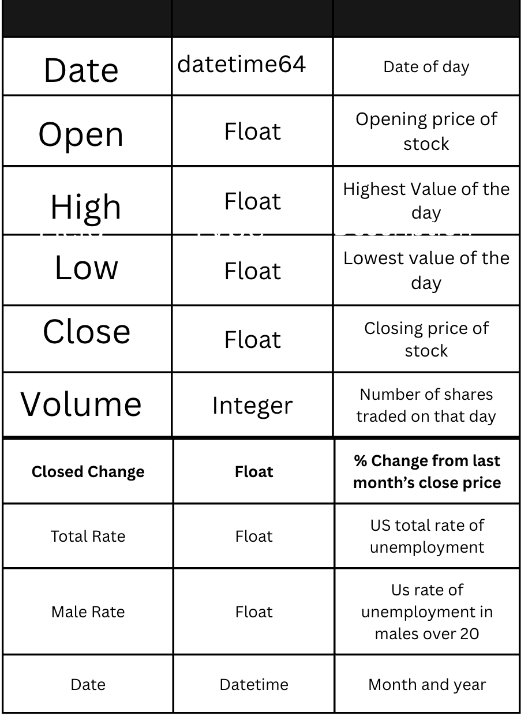
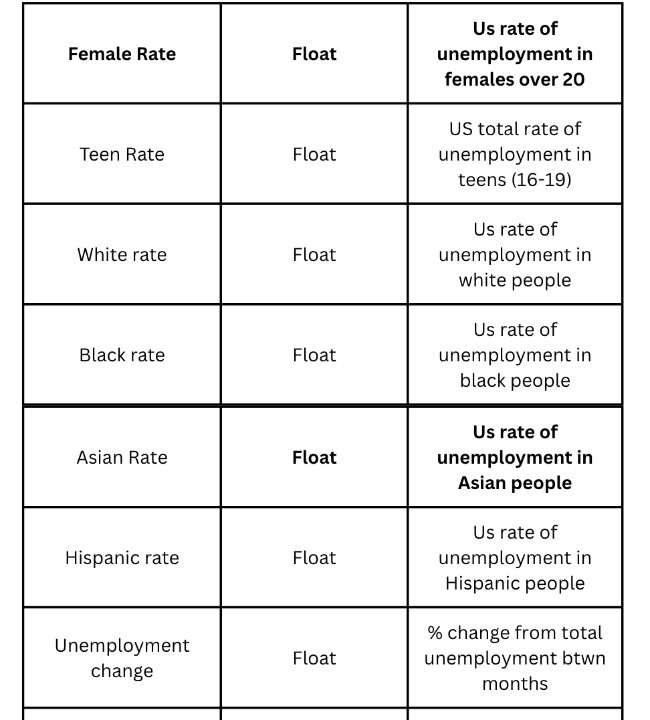
*2.2 S&P 500 Trends*

For the economic side of our analysis, we sourced a dataset from Kaggle titled "S&P 500 SPY Daily Price History🕵️", contributed by user “GK.” The dataset, referred to as spy.csv in our code, includes over 30 years of daily price data ranging from January 1993 to April 2025. With its most recent update on May 3rd, 2025, the dataset is current and free of missing values. Our cleaning process involved converting the Date column to datetime format and ensuring that all numerical variables were stored as floats. Additionally, we engineered a new column, Close Change, to represent the month-over-month change in the closing price of the S&P 500

index. The final data was put into a data frame labeled sp500\_df. It contained 8,107 observations across 11 columns.

*2.3 Combining Unemployment and S&P 500*

To perform the analysis, we executed an inner join on the Date column, which both datasets shared. We ensured consistency by separately formatting the Date column in each dataset before merging them into a master Data Frame. This merge enabled a direct, month-by-month comparison of unemployment rates and stock market performance. We ensured each feature in both data frames aligned to month-level datetime through a ‘dt.to-period’ function, and ensured all datatypes were numeric and rounded. The final data was put into a ‘merged’ data frame.

Some additional information on our data sources and preparation includes using Selenium to web scrape. We used Selenium WebDriver to open and navigate to the BLS website. To extract the table, we coded it so selenium would click “Show Table” to reveal the data and extract the unemployment rates for the various demographic groups. We stored the data for the demographics groups in lists and created a data frame.

*Table 1: Data Dictionary Data Dictionary Continued*

Source 1 (Web page): [https://www.bls.gov/charts/employment,situation/civilian,unemployment,rate.htm](https://www.bls.gov/charts/employment-situation/civilian-unemployment-rate.htm)

Source 2 (Kaggle Dataset): [https://www.kaggle.com/datasets/gkitchen/s,and,p,500,spy](https://www.kaggle.com/datasets/gkitchen/s-and-p-500-spy)

**Analysis**

*3.1 Fluctuations and Performance between S&P 500 and Unemployment Rates*

Our analysis explores whether unemployment rates across different demographic groups can serve as predictors or indicators of the S&P 500 stock market index. Unemployment data for groups including Male, Female, Teen, White, Black, Asian, and Hispanic individuals were examined alongside S&P 500 closing prices. The analysis involved descriptive statistics, bivariate regressions, hypothesis testing, and logistic regression to assess relationships and potential predictive power.

A screenshot of a graph

AI-generated content may be incorrect.The univariate analysis revealed substantial variation in unemployment levels across demographic groups. Teen unemployment had the highest mean (15.91%) and standard deviation (5.30%), highlighting the heightened vulnerability of younger workers to economic shifts. Black (mean: 9.94%, std. dev.: 3.59%) and Hispanic (mean: 7.35%, std. dev.: 3.11%) populations also faced elevated unemployment rates. In contrast, Asian (mean: 4.67%, std. dev.: 2.23%) and White (mean: 5.24%, std. dev.: 2.11%) groups exhibited lower and more stable rates.

*Table 2: Table showing results of univariate analysis for each demographic*

Female unemployment showed a skewness of 1.66 and kurtosis of 4.25, while Asian unemployment was even more extreme (skewness: 2.02, kurtosis: 5.69), indicating a greater presence of outliers. Descriptive statistics for the S&P 500 revealed a high standard deviation (138.28), moderate right skew (0.91), and elevated kurtosis (-0.34), suggesting considerable volatility and a tendency toward extreme values in market performance.

The correlation analysis reveals that changes in the S&P 500 (measured by "Close Change") have minimal association with unemployment rates across demographic groups. The strongest correlation, as expected, is with itself (1.000). Among unemployment indicators, the Asian unemployment rate shows the highest correlation with the S&P 500 at 0.0815, followed by female (0.0439), Hispanic (0.0413), and Black (0.0391) unemployment rates. Male, White, and Teen unemployment rates display even weaker correlations, all below 0.04. These low correlation values suggest that short-term movements in the stock market are largely independent of contemporaneous unemployment rates across demographic groups, highlighting the complexity and potential lag in the relationship between labor market A black screen with white text

AI-generated content may be incorrect.conditions and financial market performance.

*Table 3: The table shows correlations between S&P 500 changes and unemployment rates across all demographic groups*

To further explore these relationships, Welch’s t-test was used to compare the means of Male unemployment rates and S&P 500 highs. The test produced a p-value above 0.05, leading to a failure to reject the null hypothesis and suggesting no statistically significant difference between the two variables. In a second test, Pearson’s correlation between Female unemployment and S&P 500 highs showed a weak negative relationship, which was also not statistically significant. These results reinforce the notion that demographic unemployment rates on their own do not exhibit strong or consistent associations with stock market performance.

A graph with red dots

AI-generated content may be incorrect.*Table 4: Graph showcasing relationship between male unemployment and S&P 500 price*

A logistic regression was employed to assess whether unemployment rates could predict the daily direction of the S&P 500, using a binary outcome for market movement (1 for increase, 0 for no increase or decrease). All seven unemployment categories were used as predictors. None of the coefficients were statistically significant, and the model’s R-squared value was very low, indicating that unemployment rates lack predictive power regarding market direction.

In summary, while some unemployment rates, particularly for Teens and Black individuals, display a slightly stronger negative correlation with the S&P 500, these relationships are weak and not statistically significant. Unemployment trends alone do not adequately explain or forecast stock market movements. This underscores the complexity of financial markets, which are influenced by a broader set of factors such as interest rates, inflation, earnings expectations, and investor sentiment. Future research could benefit from integrating additional macroeconomic indicators to build more comprehensive and accurate models of market behavior.

*3.2 S&P 500 Index Increase/Decrease and its target demographic group*

This analysis explores whether changes in unemployment rates across different demographic groups can predict the short-term direction of the S&P 500 Index. The underlying idea is that labor market conditions often reflect broader economic health, and shifts in unemployment, particularly among more economically sensitive groups, might provide early signals of market movement. The hypothesis was that month over month-over-month changes in demographic-specific unemployment rates could be used to predict whether the S&P 500 would increase or decrease in a given month.

An initial visual inspection revealed no clear or consistent patterns linking any single group’s unemployment changes with stock market direction. However, major macroeconomic events such as the 2008 financial crisis and the COVID-19 pandemic caused pronounced and simultaneous movements in both unemployment and the S&P 500. These moments of

A graph of a graph

AI-generated content may be incorrect.

*Table 5: Graph showcasing change of unemployment in all demographics versus S&P 500 price*

alignment suggests that system-wide shocks, rather than changes within specific demographic groups, are more influential drivers of market behavior.

Table: The graph above shows an overview of the unemployment rate and S&P 500 closing prices

A graph with a line

AI-generated content may be incorrect.Several modeling techniques were applied to further investigate potential predictive relationships. A Random Forest classifier was trained on unemployment changes to predict S&P 500 direction. The model achieved a lackluster accuracy, around 59 percent, and had an AUC of only 0.41.

*Table 6: The graph above shows our lackluster ROC and AUC*

Feature importance scores were relatively evenly distributed across all variables, with changes in teen, Black, and Asian unemployment ranking slightly higher, though not decisively. A logistic regression model was also employed but performed poorly, offering weak predictive value and producing coefficients with little economic interpretability.

To test for potential delayed effects, a lagged variable approach was implemented, shifting unemployment changes by one month to allow for delayed market responses. This version of

the model showed a slight increase in accuracy, since it accounted for non-instant employment effects, but overall reflected the long-term upward trend in the S&P 500 rather than capturing

meaningful unemployment-market relationships. Our best metric was our True Negative A graph of a logistic regression

AI-generated content may be incorrect.rate in our confusion matrix, but we believe this to be due to countrywide economic effects.

*Table 7: The graph compares coefficients between both the lagged and original variables*

*A graph with blue and white lines

AI-generated content may be incorrect.Table 8: Feature importance including both original variables and lagged variables*

A blue and white diagram

AI-generated content may be incorrect.

*Table 9: Confusion matrix showcasing our strong True Negative rate*

Chi-square tests of independence were also conducted on binned unemployment change variables to assess statistical association, but none of the demographic groups demonstrated significant relationships with market direction.

Overall, the analysis does not support the hypothesis that changes in demographic-specific unemployment rates can predict movements in the S&P 500. Across all methods tested, the models failed to deliver consistent, meaningful, or interpretable results. While unemployment rates are undeniably important economic indicators, their predictive value for short-term stock market movements appears limited when disaggregated by demographic group. Instead, market behavior seems to respond more strongly to broader, system-wide economic conditions. Future research in this area may be better served by focusing on comprehensive macroeconomic indicators, such as inflation rates, interest rate changes, or consumer sentiment, that more directly reflect market-relevant economic dynamics.

*3.3 Model Accuracy based on large economic event*

**A diagram of a map

AI-generated content may be incorrect.**The data presented across the charts and statistical summaries highlight an important observation: while there appears to be some level of statistical significance in the relationship between demographic unemployment rates and economic indicators, there is no consistent or direct correlation between unemployment rates and S&P 500 performance across all groups. The sharp rise in unemployment rates during the 2008 recession, as seen across male, female, and racial groups, reflects a systemic economic collapse rather than market-specific behavior. Each demographic showed increases in unemployment, but these movements followed the broader downturn caused by the collapse of financial institutions and housing markets, not by fluctuations in equity markets themselves.

*Table 10: Scatter plot between unemployment rate and S&P performance during recession*

The S&P 500 vs. Unemployment scatterplot further emphasizes this disconnect. Rather than showing a linear or strongly correlated trend, the data points are widely dispersed, indicating that high or low unemployment rates did not reliably align with specific levels of the S&P 500 index. This dispersion undermines any claim of a direct relationship. The S&P 500 is forward-looking and often buoyed by investor sentiment, monetary policy, and corporate earnings expectations, while unemployment is a lagging indicator that reflects deeper-rooted labor market responses to past economic conditions. This difference in behavior contributes to their unaligned patterns.

When considering the 2008 financial crisis specifically, both unemployment and the S&P 500 were responding to the same underlying cause: a profound collapse in the U.S. economic structure, fueled by risky mortgage lending, financial deregulation, and banking failures. This shared origin explains why both metrics reacted dramatically, but it does not mean they are

causally linked to one another. Instead, the economic event itself acted as the driver of both unemployment spikes and market declines. It is more accurate to say that large-scale economic

**A graph of different colored lines

AI-generated content may be incorrect.**

*Table 11: Showcasing spikes in unemployment during the 2008 recession*

disruptions steer both metrics in the same direction, rather than claiming one reliably predicts or causes the other.

The t, t-statistics and p, p-values associated with unemployment data across demographic groups confirm that changes were statistically significant, especially during the recession period. However, statistical significance here reflects how sharply unemployment rose during a crisis, it doesn’t imply causation with S&P 500 trends.

In truth, both unemployment and equity market performance are co-moved by macroeconomic health. Pandemics, recessions, or geopolitical instability influence consumer behavior, investor confidence, business investment, and job creation all at once. Therefore, the most accurate model recognizes that the health of the American economy is the root driver behind these seemingly related indicators, not the indicators themselves driving each other.

**4. Conclusion**

In this project, we analyzed three core aspects of unemployment rates and economic market trends: the overall correlation between unemployment across demographic groups and the S&P 500 index, how specific demographics respond to rising or falling of the S&P 500 Index, and  
how these relationships change during major economic events such as the 2008 recession.

In summary, based on the analysis questions presented in our proposal, we found the following results:

1. **How do fluctuations in unemployment rates across various demographic groups correlate with the performance of the S&P 500 index?**

Our analysis revealed a limited correlation between unemployment in various demographic groups and the average performance of the S&P 500.

1. **If the S&P 500 is increasing or decreasing, which demographic is most closely related to the S&P performance?**

There is moderate correlation between a declining S&P 500 and increases in unemployment among teenagers. However, there is little observable correlation between S&P 500 increases and changes in any particular demographic group’s unemployment rate.

1. **Honing specifically on a large economic event (e.g., the 2008 recession), is our model more accurate?**

During the 2008 recession, the correlation between the S&P 500 and unemployment rates among demographic groups strengthened. However, both trends were likely responding to a larger macroeconomic force (the recession itself) rather than influencing one another directly.

This project faced a few limitations, including the absence of additional economic indicators (e.g., interest rates, inflation, consumer spending, other macroeconomic events) that could provide a more complete picture. The inclusion of such variables could strengthen correlations and improve the robustness of the models. Future work should focus on incorporating 3–4 additional datasets (either scraped from reputable sources or downloaded) to enhance predictive power and contextual depth.

Although our models often produced weak correlations between unemployment and S&P 500 performance, this is a meaningful insight. It suggests that while both the labor market and the stock market are central to the economy, they tend to operate on different timescales and respond to different forces. Therefore, policymakers and investors should be cautious in assuming a strong direct relationship between the two. Understanding this distinction can lead to more informed economic decisions and highlight the importance of analyzing multiple variables when evaluating national economic health.