Unemployment Rate In United States Statistical Analysis

by

Ishaan Bandekar, Shivanshu Vinay Singh, Yash Karbhari, Jayant Bangia

FINAL REPORT

Submitted to the Faculty of the Stevens Institute of Technology
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE - DATA SCIENCE

Advisor: Prof. Hong Do

STEVENS INSTITUTE OF TECHNOLOGY
Castle Point on Hudson
Hoboken, NJ 07030
2024

Table of Contents

1.0 Introduction

1.1 About the Unemployment Rate in the USA:

The unemployment rate represents the number of unemployed as a percentage of the labor force. Labor force data are restricted to people 16 years of age and older, who currently reside in 1 of the 50 states or the District of Columbia, who do not reside in institutions (e.g., penal and mental facilities, homes for the aged), and who are not on active duty in the Armed Forces.

1.2 Data Description:

Total labor force: https://fred.stlouisfed.org/series/UNRATE
Asian: https://fred.stlouisfed.org/series/LNU04032183
Black: https://fred.stlouisfed.org/series/LNS14000006
Hispanic: https://fred.stlouisfed.org/series/LNS14000009
Men: https://fred.stlouisfed.org/series/LNS14000001
Women: https://fred.stlouisfed.org/series/LNS14000002
White: https://fred.stlouisfed.org/series/LNS14000003
We have a total of 253 rows, each comprising of 7 features which are,

- Unemployment rate of the total labor force
- Unemployment rate of Asian
- Unemployment rate of Black
- Unemployment rate of Hispanic
- Unemployment rate of White
- Unemployment Rate of Men
- Unemployment rate of Women

1.3 Goal of the project:

Our project aims to develop a robust regression model for predicting future unemployment rates in the USA, focusing on demographic factors such as gender and ethnicity. By leveraging a combination of descriptive statistics, inferential analysis, and data visualization techniques, we seek to uncover valuable insights into the complex dynamics of unemployment. The primary objectives of our study include:

- Investigating the relationship between unemployment rates and demographic variables, particularly gender and ethnicity.
- Identifying demographic groups with the most favorable job prospects.
- Exploring potential factors or events contributing to observed trends in unemployment rates across various demographic segments.
- Developing a predictive model to forecast future unemployment rates for different demographic groups, facilitating informed decision-making and policy planning.

2.0 Methods

Our initial focus will be on data visualization, where we'll create graphical representations to explore the relationships between demographic factors and unemployment rates. Following this, we'll conduct descriptive analysis using statistical measures to summarize the central tendencies and variabilities within the dataset, leading to valuable insights.
Furthermore, inferential statistics will be employed to investigate the correlation between demographic features and the target variable, which in this case is unemployment rates. Hypothesis tests will be conducted to assess specific observations derived from the relationship between certain demographic factors and unemployment rates.
Subsequently, predictive regression models will be developed. Initially, each demographic feature will be considered as input, with unemployment rates (the target variable) as the output. Subsequently, all demographic features will be utilized as input, with unemployment rates as the output variable. The performance of these regression models will be evaluated using metrics to ensure reliability and accuracy, facilitating the identification of the best-performing model.

3.0 Statistical Methods

3.1 Descriptive Analysis:

	DATE	unrate	unrate_asian	unrate_black	unrate_hispanic	unrate_men	unrate_white	unrate_women
count	253	253.000000	253.000000	253.000000	253.000000	253.000000	253.000000	253.000000
mean	2013-07-01 12:54:04.268774656	5.888142	4.688538	10.122530	7.354941	6.079842	5.224111	5.673123
min	2003-01-01 00:00:00	3.400000	2.000000	4.800000	3.900000	3.400000	3.000000	3.300000
25%	2008-04-01 00:00:00	4.400000	3.100000	7.700000	5.100000	4.400000	3.900000	4.300000
50%	2013-07-01 00:00:00	5.300000	4.100000	9.700000	6.600000	5.400000	4.500000	5.200000
75%	2018-10-01 00:00:00	7.200000	5.900000	12.300000	9.000000	7.600000	6.300000	6.700000
max	2024-01-01 00:00:00	14.800000	15.000000	16.900000	18.900000	13.500000	14.200000	16.200000
std	NaN	2.052394	2.019190	3.250666	2.804140	2.211933	1.900212	1.920542

Figure 1. Descriptive Statistics of Unemployment Rates by Demographic Group

The statistical overview in Figure 1 captures the disparities in unemployment rates among different demographics. With an overall average unemployment rate of 5.88%, there is a noticeable variance among groups: Asians report a lower mean of 4.69%, while African Americans experience the highest at 10.12%. The median figures present a similar trend, with the median for African Americans at 9.7%, significantly higher than the overall median of 5.3%. The spread of the data, as shown by the standard deviation, indicates the greatest variability within the African American group at 3.25, compared to a lower variability for women at 1.92. These numbers highlight the distinct economic challenges faced by each demographic, underlining the need for nuanced economic interventions.

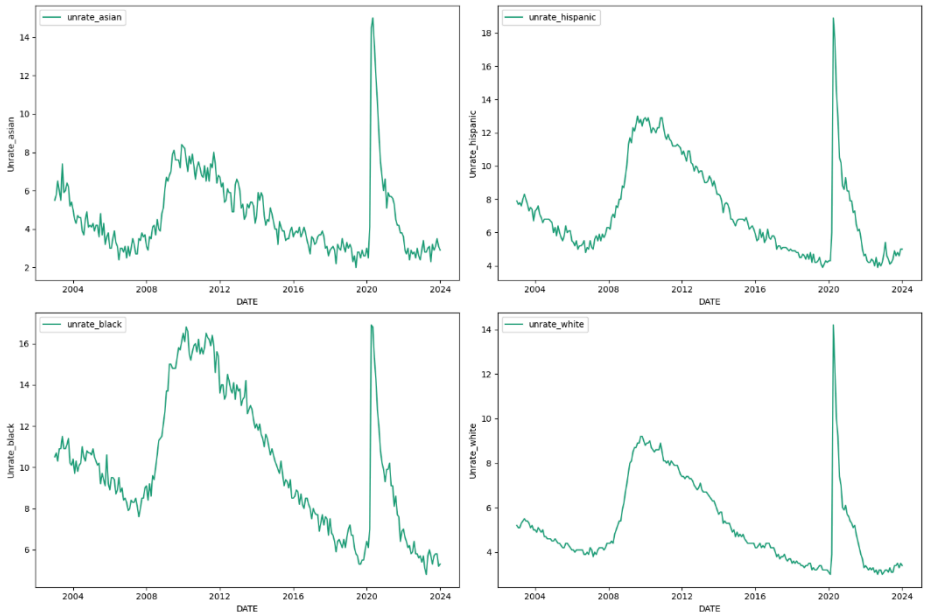


Figure 2. Ethnic Unemployment Trends in the U.S. from 2004 to 2024

Figure 2 displays four separate line graphs, each depicting the unemployment rate trends over time from 2004 to 2024 for different ethnic groups within the United States: Asian, Black, Hispanic, and White populations. These graphs are critical tools for understanding how unemployment rates have varied for each group across the span of two decades, reflecting the unique economic challenges and resilience patterns among these demographics. Such visual representations are instrumental in our descriptive analysis, allowing us to observe trends, spikes, and recovery periods specific to each ethnic group. The data illustrate variable unemployment rates among the groups with notable spikes during two critical economic crises. The Great Recession, around 2009, saw a significant increase in unemployment rates, with the Black population reaching an unemployment rate peak of approximately 16%, while the Asian community peaked at a slightly lower rate, though still substantial. The second major spike in unemployment occurred in 2020, driven by the COVID-19 pandemic, with the Hispanic community seeing a dramatic surge to nearly 18%, reflecting the severe impact of the pandemic on this demographic. Both peaks were followed by a decline in unemployment rates, suggesting a period of recovery post-crisis. For instance, the Asian community experienced a sharp decline after the 2020 peak, returning closer to pre-crisis levels. Meanwhile, the unemployment rates for White Americans peaked at around 14% during the same period, before decreasing again. These specific numbers are crucial as they not only quantify the impact of economic disruptions on employment but also show the disparate effects on different ethnic communities, emphasizing the need for tailored economic policies and support systems.

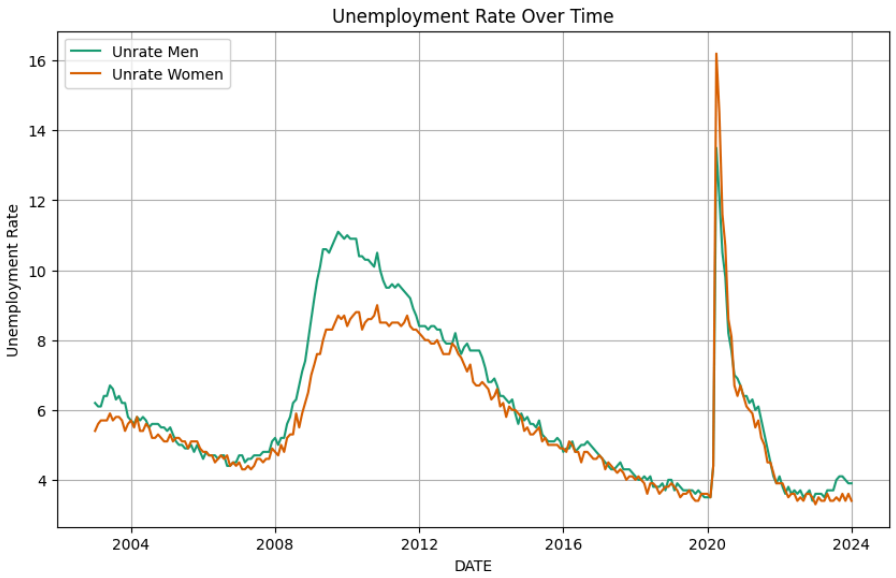


Figure 3. Gender Differences in Unemployment Rates: 2004-2024

The line graph in Figure 3 provides a comparative analysis of the unemployment rates for men and women in the United States from 2004 to 2024. It illustrates the fluctuations in joblessness among genders and underscores the effects of economic crises on employment. Notably, in 2008, the graph indicates a significant divergence between the unemployment rates of men and women, which can be largely attributed to the collapse of the housing market bubble. This economic downturn severely impacted industries that predominantly employed men, particularly those sectors that offered high-paying jobs. As a result, the unemployment rate for men escalated more sharply compared to women, depicting a gendered disparity in job loss during the recession. This point of divergence highlights the vulnerability of occupations heavily occupied by men to the housing market dynamics and the subsequent ripple effects on employment. In stark contrast, the graph reflects a role reversal during the COVID-19 crisis in 2020, with women experiencing a steeper rise in unemployment rates than men. This shift can be traced back to the concentration of women in lower-paying service-oriented sectors, such as hospitality and retail, which were hit hard by the pandemic's economic disruptions. As businesses closed and services were reduced, women faced higher job losses, evidenced by the peak unemployment rate for women surpassing that of men. The pandemic's impact reiterates the fragility of sectors where women are overrepresented and highlights the broader implications of wage and job security disparities in the workforce. These examples not only illuminate the dynamics of gender disparities in labor market outcomes but also underscore the importance of considering sectoral employment composition and wage levels when addressing the impacts of economic upheavals on different segments of the population.

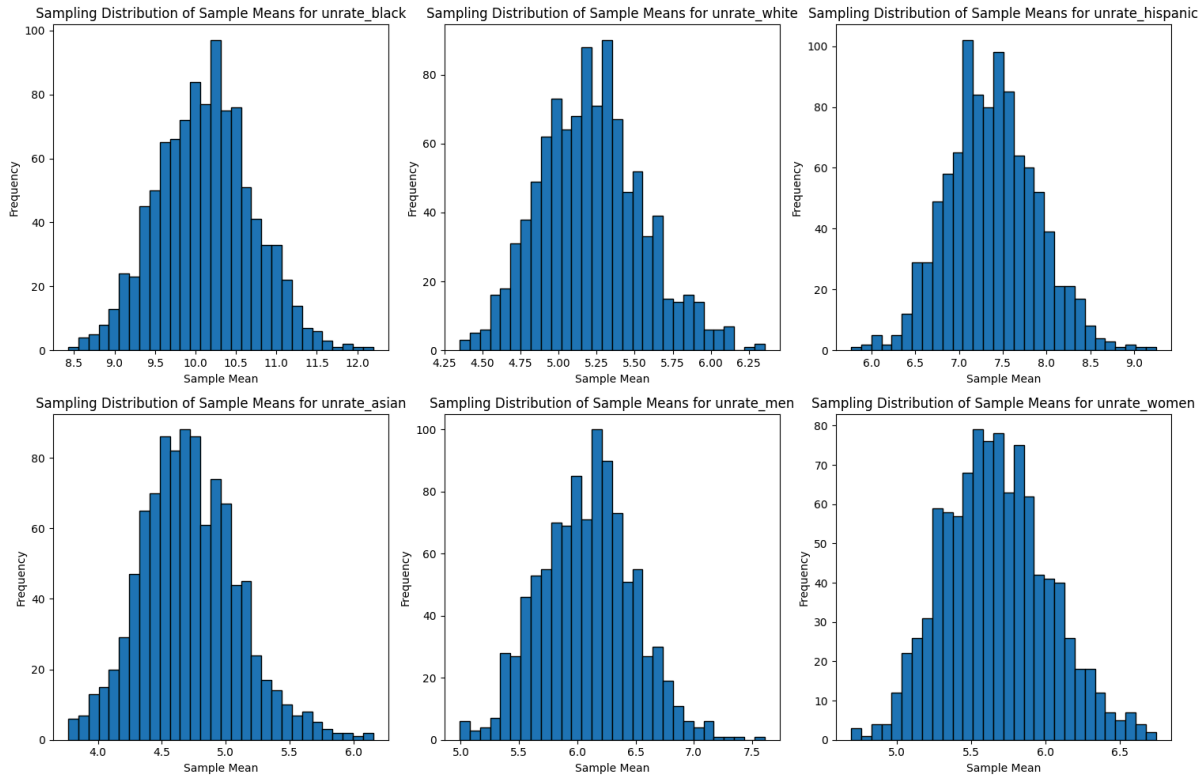


Figure 4.1 Frequency Central Distributions of Unemployment Sample Means by Demographic Group

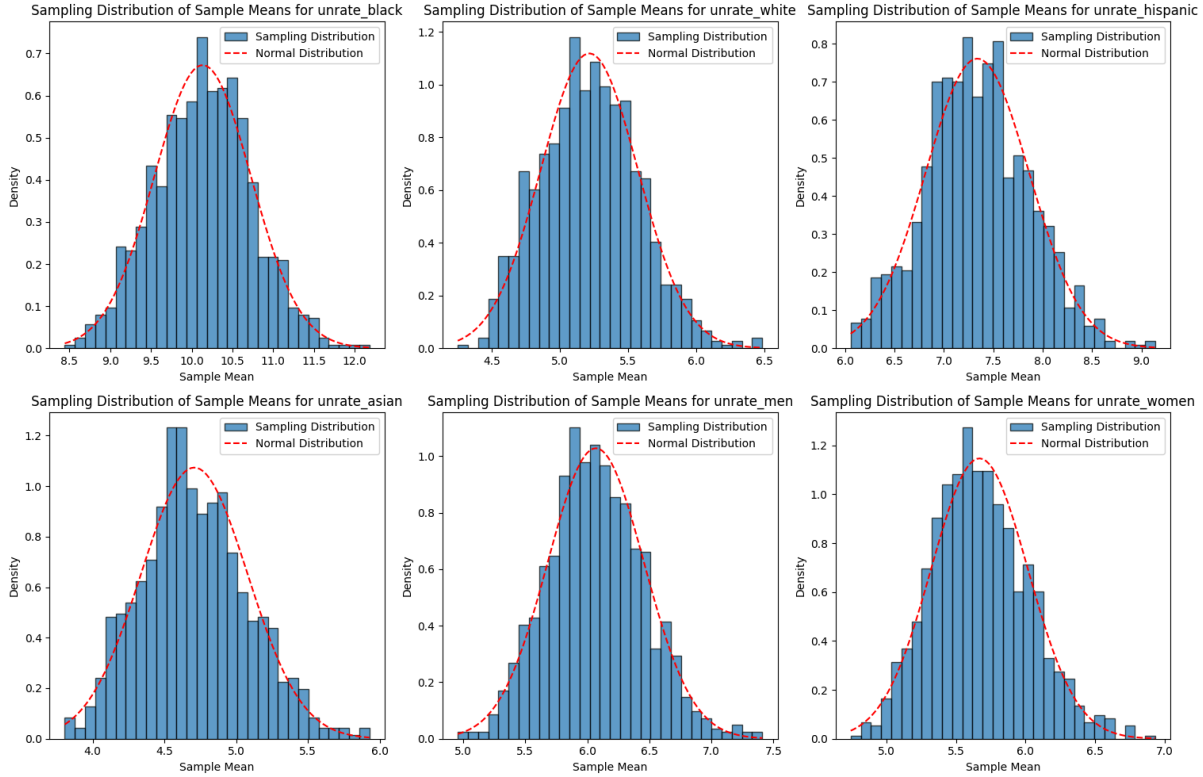


Figure 4.2 Frequency Normal Distributions of Unemployment Sample Means by Demographic Group

The image showcases six histograms illustrating the sampling distributions of sample means for unemployment rates categorized by different demographic groups: black, white, Hispanic, Asian, men, and women. Each histogram is accompanied by a dashed line representing the normal distribution curve, highlighting the application of the Central Limit Theorem (CLT). The distribution of sample means for black, white, and men's groups is broader and exhibits

higher mean values in comparison to those for Hispanic, Asian, and women's groups. Observations can be made about the centering and spread of these distributions: black, white, and Hispanic groups have higher average unemployment rates, reflected in their means of approximately 9.5, 5.5, and 7.5 respectively. This variation underscores distinct unemployment rate disparities among these demographics. These histograms demonstrate a strong alignment with their respective normal distribution curves, which validates the theoretical expectation of the CLT that sample means will tend towards a normal distribution as sample sizes increase. The close fit suggests that the sample sizes are adequate to assume normality in the sampling distributions. This visualization not only provides statistical insight into the unemployment rate dynamics across different populations but also serves as an effective demonstration of the CLT in real-world data applications, confirming the robustness of the statistical method in diverse settings.

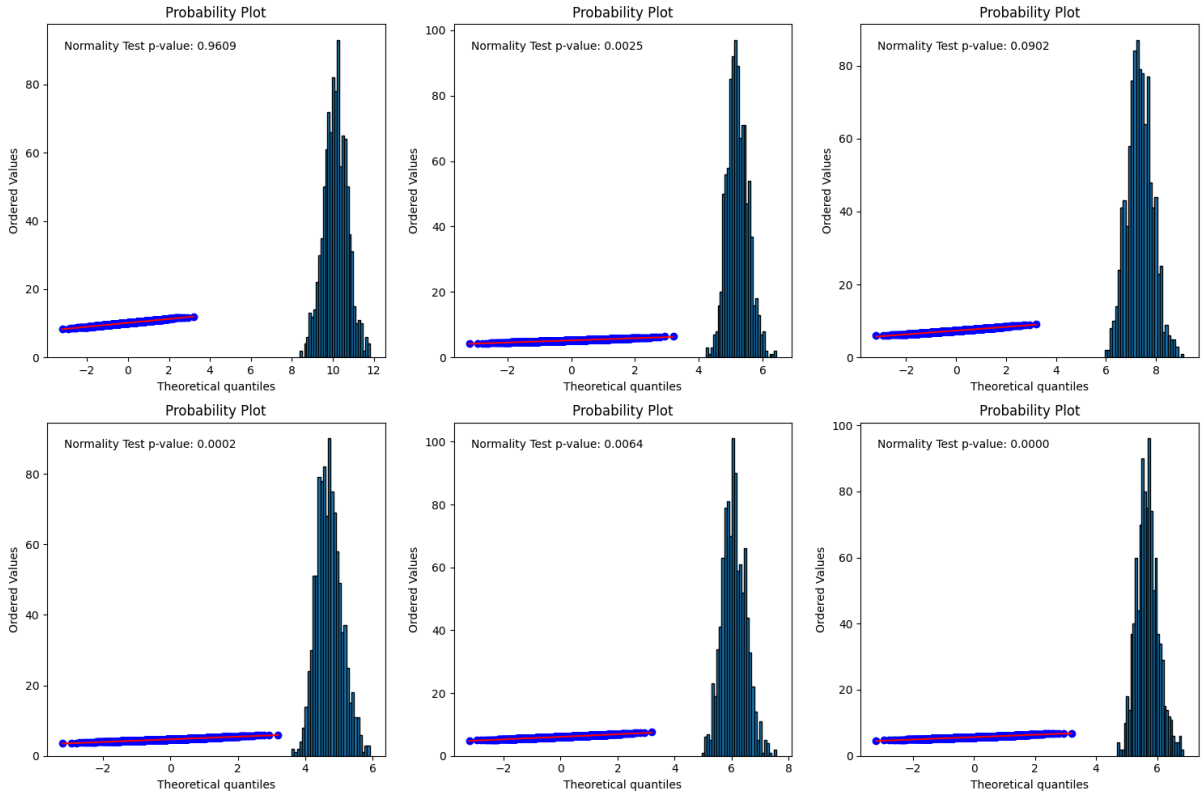


Figure 5. Q-Q plots to show normality test for each variable

The image features six probability plots, each containing a histogram and a Q-Q plot, used to evaluate the normality of datasets through normality test p-values. While the first plot indicates a normal distribution with a p-value of 0.9609, the subsequent plots, especially those with p-values of 0.0025, 0.0902, 0.0002, 0.0064, and 0.0000, show significant deviations from normality, with some datasets exhibiting extreme kurtosis and sharp peaks in histograms. Based on the visual assessment of the QQ plots, it appears that the data for unrate_black, unrate_white, and unrate_men may not be normally distributed. In these plots, the points deviate from the diagonal line, which suggests that the data is skewed. The p-values for these tests are also very low (less than 0.05), which statistically supports the conclusion that the data is not normally distributed.

On the other hand, the QQ plots for unrate_hispanic, unrate_asian, and unrate_women appear to be approximately linear, and the p-values for these tests are all greater than 0.05. This suggests that the data for these columns may be closer to a normal distribution. It's important to note that normality tests are sensitive to sample size. So, if your sample size is small, the test may not be reliable. However, the visual assessment of the QQ plots is a more general approach that can be useful with any sample size.

Overall, the data from this analysis appears to be mixed. Some columns show signs of normality, while others do not. This may be an important consideration depending on the specific statistical methods we plan to use in our project.

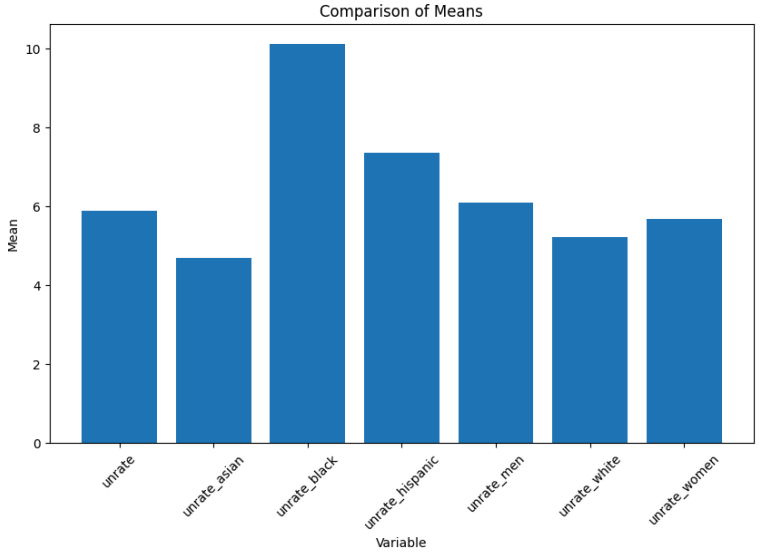


Figure 6. Frequency Distributions of Unemployment Rates

The bar graph in Figure 6 offers a clear comparison of the mean unemployment rates for different demographic groups, categorized by ethnicity and gender. This straightforward visual tool is crucial for quickly identifying disparities in unemployment across these varied groups within a given timeframe. The bars represent the average unemployment rates, making it easy to see which groups experience higher or lower rates relative to each other. Upon reviewing the graph, we can see that the mean unemployment rate is highest among the Black population at approximately 9, followed by the Hispanic group with a mean of around 8. The Asian and White populations exhibit lower mean unemployment rates, with figures near 6 and 7 respectively. When comparing gender, the mean unemployment rate for men is marginally higher than for women, with men's rates close to 7 and women's around 6.5. These figures are critical as they not only reflect the unemployment scenario within each group but also highlight the need for targeted economic policies to address the disparities. Understanding these numbers helps in assessing the effectiveness of past interventions and planning future strategies to improve employment opportunities for these demographic segments.

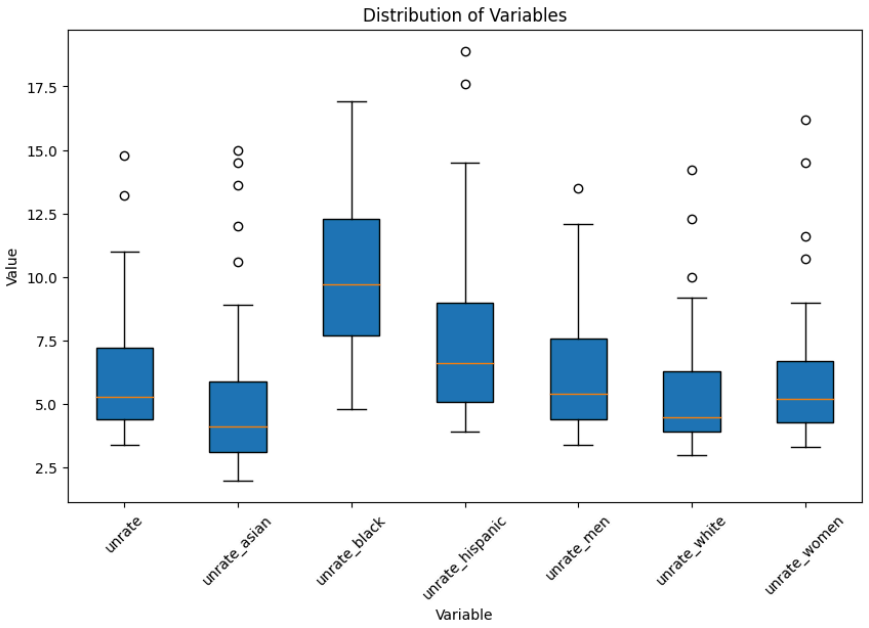


Figure 7. Unemployment Rate Distribution Across Demographic Groups Boxplot

The box plot in Figure 7 presents the distribution of unemployment rates for seven distinct demographic groups: the general population, Asian, Black, Hispanic, men, White, and women. Each box plot highlights the median of the dataset with a bold horizontal line, the interquartile range (IQR) with the box edges, and potential outliers with individual dots. Outliers are data points that fall beyond the whiskers, which extend to 1.5 times the IQR from the quartile boundaries. The plot reveals differences in the central tendency and dispersion of unemployment rates among the groups. The median unemployment rates for these groups span from around 5 to over 10, indicating a wide variance in unemployment experiences. The Black demographic shows a notably higher median, suggesting a greater central unemployment rate, while Asian and White groups appear to have lower median rates, indicative of lower central unemployment trends within these groups. The plots for men and women display medians closer to each other, reflecting more similarity in their central tendencies. The presence of several outliers in groups like the Black and Hispanic demographics points to exceptional cases that stand out from the common trends and might require additional scrutiny. The information gleaned from these distributions is critical for identifying groups that may experience higher variability in unemployment and for guiding further analytical or policy-oriented inquiries.

3.2 Inferential Analysis:

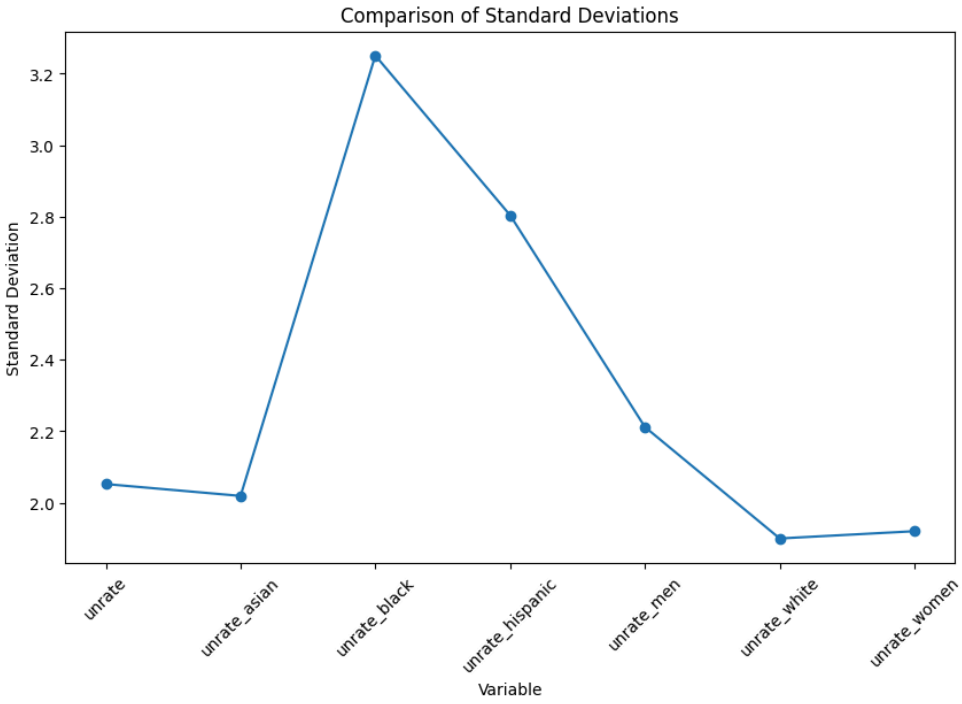


Figure 8. Variability in Unemployment Rates Among Demographics

The line chart in Figure 8 here plots the standard deviation of unemployment rates for different demographic groups, which provides insights into the variability or dispersion of unemployment figures within each group. Standard deviation is a key statistical measure that indicates how much the individual data points differ from the mean value of the dataset. From the chart, we observe that the standard deviation for the unemployment rate of the Black demographic peaks the highest among all groups, showing a value slightly above 3. This indicates a high level of variation in unemployment rates within this group. In contrast, both the Asian and Women demographics exhibit the lowest variability, with their standard deviations hovering around the 2.2 mark. The Hispanic group also shows a considerable level of variation, with a standard deviation approaching 3, just slightly less than that of the Black demographic. The standard deviations for the Men and White groups show moderate variability, with values around 2.5 and 2.3 respectively. These numbers are particularly useful in understanding how the spread of unemployment rates within each group compares, indicating which groups experience more consistency in employment levels and which groups face more fluctuation.

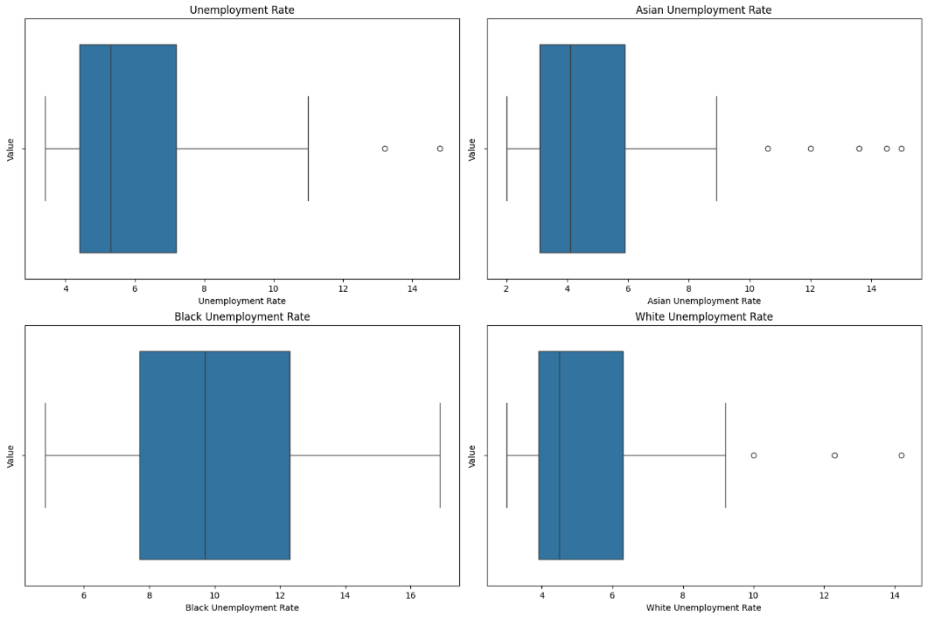


Figure 9. Comparative Unemployment Rates by Race and Overall

The collection of box plots in Figure 9 provides a comparison of unemployment rates across different racial groups. These plots display the range, median, interquartile range (IQR), and outliers of unemployment rates for the general population and Asian, Black, and White groups. The median is indicated by the line within each box, the IQR is represented by the box itself, and the 'whiskers' extend to the furthest points within 1.5 times the IQR from the quartiles. Analyzing the plots, it's evident that the median unemployment rate for the Black group is the highest, significantly greater than the others, showing a central tendency above 10. In contrast, the Asian unemployment rates have a lower median, suggesting a tighter concentration of data around a lower central tendency. The general and White unemployment rates show similar distributions, with medians slightly below 6. Outliers are present in each group, indicating individuals with unemployment rates notably different from the bulk of the data. These points are essential for understanding the spread and central tendency of unemployment within these racial categories and for identifying rates that deviate significantly from the majority, potentially signaling the need for targeted economic or employment policies.

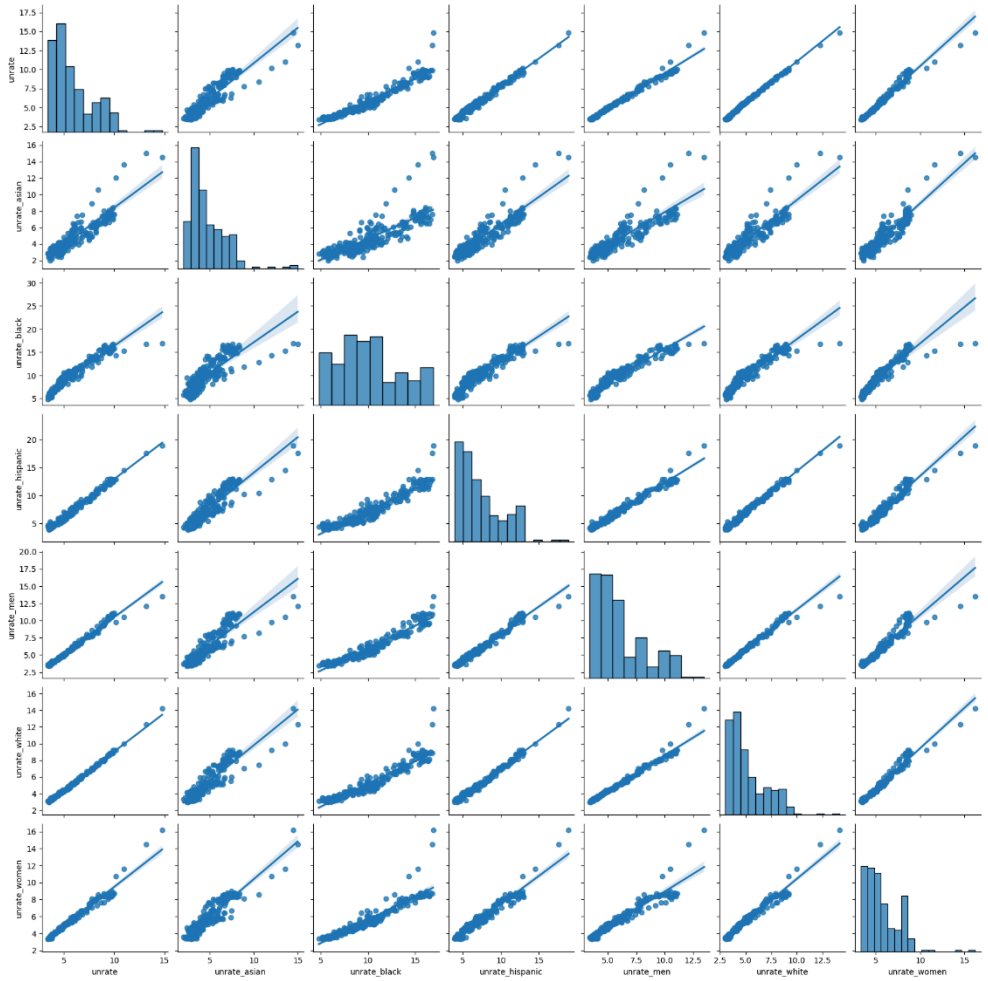


Figure 10. Correlations and Distributions of Unemployment Across Demographics

Figure 10 is a matrix of scatter plots, combined with histograms along the diagonal, showcasing the relationships between various unemployment variables. Each scatter plot compares two variables, illustrating the degree and direction of their relationship, while the histograms display the distribution of single variables. The matrix enables the identification of trends and potential correlations between different demographic unemployment rates. Across the matrix, the scatter plots with upward trends suggest positive correlations between the variables they compare, meaning as one variable increases, the other tends to increase as well. Conversely, a downward trend would indicate a negative correlation. However, all the visible scatter plots in this matrix show positive correlations. The histograms reveal the distribution of each variable, with most showing a central peak and a symmetric spread, indicative of normally distributed data. This correlation matrix is a comprehensive tool for identifying and visualizing relationships between different sets of unemployment data.

3.2.1 Heat Map:

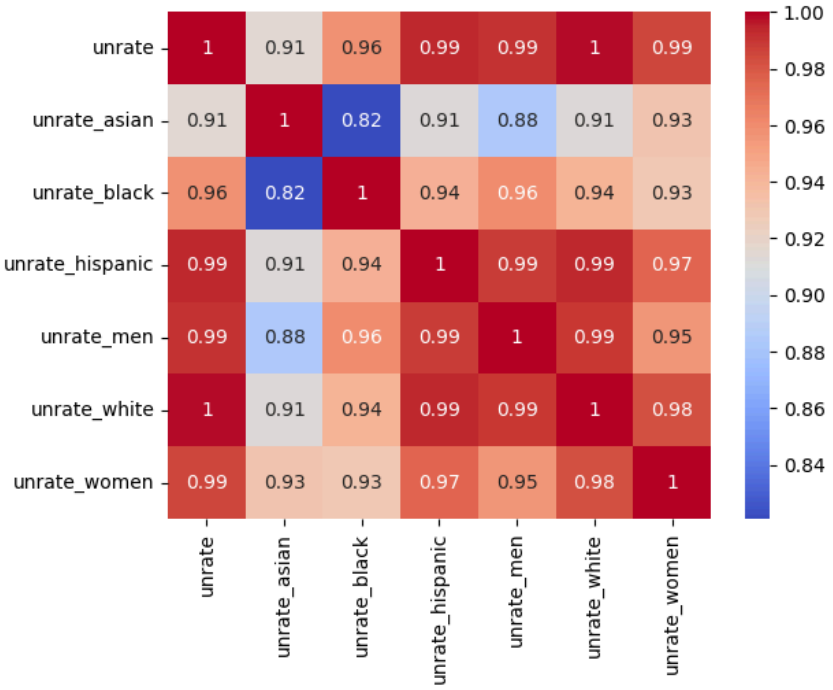


Figure 11. Correlation Heatmap of Unemployment Rates by Demographic

This heatmap in Figure 11 visually captures the correlation coefficients between unemployment rates of different demographic groups. The color intensity reflects the strength of the relationship, with red indicating a higher correlation. The general unemployment rate is highly correlated with all other categories, especially with the Hispanic and men's unemployment rates, both registering correlations nearly at unity. These strong correlations suggest that the unemployment trends for these groups closely mirror the overall unemployment trend. In contrast, the correlation between the unemployment rates for Asians and African Americans is less pronounced at 0.82, which might suggest that unique economic factors or conditions impact these groups differently. Despite this, the correlations across the board are notably strong, all above 0.90, except for this one instance, implying that while the magnitude of unemployment rates may vary among groups, the fluctuations and changes over time tend to be similar. This consistent pattern indicates a degree of systemic influence affecting unemployment rates across demographics, which is crucial for policymakers to consider when designing targeted economic interventions.

3.2.2 Hypothesis Testing:

Hypothesis Testing

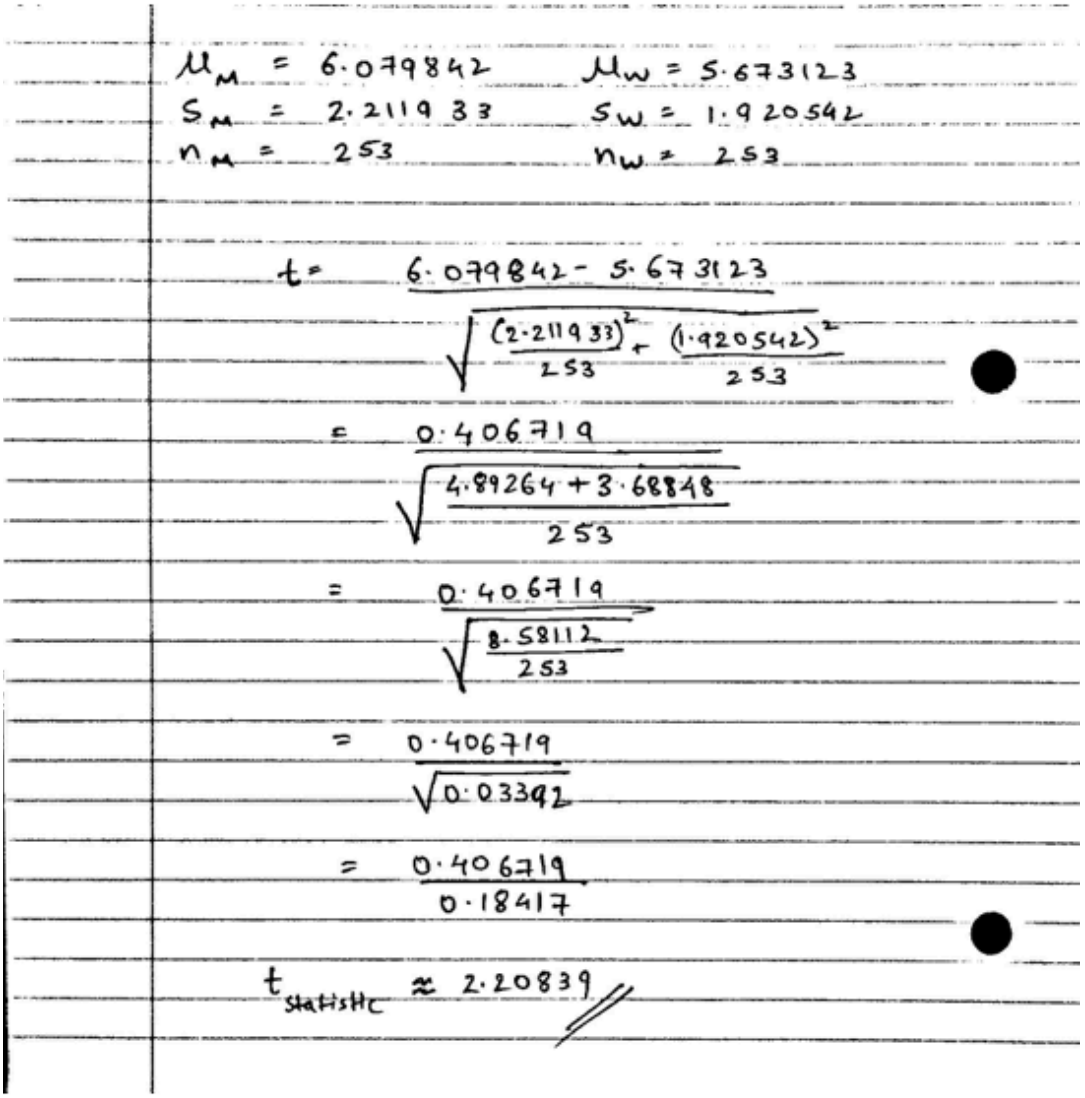
2 Tailed test:

Independent 2-sample t-test is appropriate since we have two independent groups (in this case, female and male workers) and we want to compare the means of a continuous variable (in this case, the general unemployment rates). We therefore performed this independent t-test to check and confirm the hypothesis whether male workers actually are more unemployed than female workers as we have observed.

- Null Hypothesis (H_0): There is no difference in the mean unemployment rate between male and female workers.

- Alternative Hypothesis (H_1): Male workers are more unemployed than female workers.
- Significance Level (α): 0.05. (which represents a 5% chance of rejecting the null hypothesis when it is true)
- Number of Male workers in the dataset = 253.
- Number of Female workers in the dataset = 253.
- Degree of freedom (DOF) = Male + Female - 2 = 504.
- T-critical [$T_{504}(\alpha=0.05)$] = 1.6479. (source: https://www.ttable.org/student-t-value-calculator.html#google_vignette).

$$t = \frac{(\bar{X}_M - \bar{X}_W)}{\sqrt{\left(\frac{s_M^2}{n_M} + \frac{s_W^2}{n_W}\right)}}$$



The t-test produces a t-statistic and a p-value. The t-statistic measures the difference between the group means relative to the variability within the groups. The p-value indicates the probability of observing such extreme results if the null hypothesis is true.

The decision rule is to compare the T-statistic to the T-critical (from t-table) are:

- If the T-statistic is greater than T-critical, you reject the null hypothesis.
- If the T-statistic is less than T-critical, you fail to reject the null hypothesis.

The decision rule is to compare the p-value to the significance level (alpha):

- If the p-value is less than alpha, you reject the null hypothesis.
- If the p-value is greater than or equal to alpha, you fail to reject the null hypothesis.

T statistics = 2.20839 > T critical(1.6479) → REJECT Ho
P-value: 0.013836 < 0.05 → REJECT Ho

Therefore, based on the dataset used in this study, we can confirm our hypothesis that MALE WORKERS ARE MORE UNEMPLOYED THAN FEMALE WORKERS.

Z-Test:

We conducted a hypothesis test using a z-test to investigate whether male workers experience a higher unemployment rate compared to female workers. The purpose of this analysis was to determine if there is evidence to support the claim that one gender group faces a significantly higher level of unemployment than the other. To perform the test, we collected data on the unemployment rates of male and female workers, represented by the columns 'unrate_men' and 'unrate_women' respectively.

We formulated our hypotheses as follows:

Null hypothesis (H0): There is no difference in unemployment rates between male and female workers.

Alternate hypothesis (H1): Male workers have a higher unemployment rate than female workers.

To evaluate these hypotheses, we calculated the mean and standard deviation of the unemployment rates for both male and female workers. Subsequently, we applied a one-sample z-test, comparing the unemployment rate of male workers against the mean unemployment rate of female workers under the assumption of equal population standard deviation. The z-test yielded a z-statistic and a corresponding p-value.

Interpreting the results, if the obtained p-value is less than our chosen significance level (typically 0.05), we reject the null hypothesis in favor of the alternate hypothesis. This indicates that there is sufficient statistical evidence to conclude that male workers indeed experience a higher unemployment rate than female workers. Conversely, if the p-value is greater than or equal to the significance level, we fail to reject the null hypothesis, suggesting

that there is not enough evidence to support the claim of a significant difference in unemployment rates between male and female workers.

Table 3.2.2 Hypothesis testing Results

```
Z-statistic: 3.3751363637679934
P-value: 0.00036889561632946766
Reject the null hypothesis.
Since P-value is less than Significance level
There is sufficient evidence to suggest that male workers have a higher unemployment rate than female workers.
```

Conclusion: There is enough evidence to suggest male workers indeed experience a higher unemployment rate than female workers.

4.0 Prediction Model Building

4.1 Linear Regression:

First, we performed simple linear regression, utilizing the Linear Regression Algorithm, using each of the individual features as input and target (unrate) as output. The correlation, coefficient, standard error, and p-value for each regression are summarized in table (table no.1) below. We can observe that the “unrate_white” has the highest correlation coefficient. This suggests that changes in the unemployment rate are closely related to changes in the unemployment rate among the white population as compared to other races. Also, The lowest p-value for the "unrate_white" feature suggests that variations in the unemployment rate among the white population are particularly significant in predicting overall unemployment rates.

Table no. 1: Summary of Simple Linear Regression Analyses performed using the Individual Features (as input)

Feature	Correlation	Coefficient	Standard Error	P-value
unrate_men	0.9920219679582721	0.4255505400583283	0.027585753495497734	1.2815204773492242e-35
unrate_women	0.9859093654507795	0.3905544890075142	0.022230706678514914	4.816157170572885e-42
unrate_white	0.9980563580840371	0.1457904664724542	0.041857003935127736	0.0006119411364380394
unrate_black	0.9591261378902814	0.019453911452252448	0.005954164346301595	0.0012831179921868612
unrate_asian	0.9139519819008098	0.0038480462739604662	0.004448377263860513	0.38807705935207115
unrate_hispanic	0.9930349642515867	0.012692943058088518	0.008097359829919056	0.11860960175894282

4.2 Multiple Linear Regression:

Next, we performed a multiple linear regression, utilizing the Linear Regression Algorithm. We performed two Regressions firstly for the features with races as input in the first regression and secondly using features with gender as input and target (unrate) as output. The coefficient, intercept, standard error, and p-value for both the multiple regression Analyses are summarized in the tables below.

4.2.1 Multiple Regression Based on Races:

Table no. 2: Summary of Multiple Linear Regression Analyses performed for features with races as input.

Feature	Coefficient	Standard Error	P-value
Intercept	0.09855387997910126	0.014218673028058442	5.80765945439012e-11
unrate_white	0.8247078114535469	0.01813416554345014	1.9929920022089285e-106
unrate_black	0.10849343381570326	0.003961063096532214	4.2573501653816024e-69
unrate_asian	0.05847661540671351	0.005100294550933888	1.2077265805297107e-23
unrate_hispanic	0.01470227393011489	0.012536362804066396	0.24230407421008837

Multiple Linear Regression Prediction Equation for features with races:
unrate = 0.10 + (0.82 * unrate_white) + (0.11 * unrate_black) + (0.06 * unrate_asian) + (0.01 * unrate_hispanic)

4.2.2 Multiple Regression Based on Gender:

Table no. 3: Summary of Multiple Linear Regression Analyses performed for features with gender as input.

Feature	Coefficient	Standard Error	P-value
Intercept	0.0011825896001811664	0.008234587426804902	0.8859517986437063
unrate_women	0.46744801260444185	0.0046725327800021775	4.029879976438028e-172
unrate_men	0.532074954826441	0.0039910236012243765	1.4331638251713315e-196

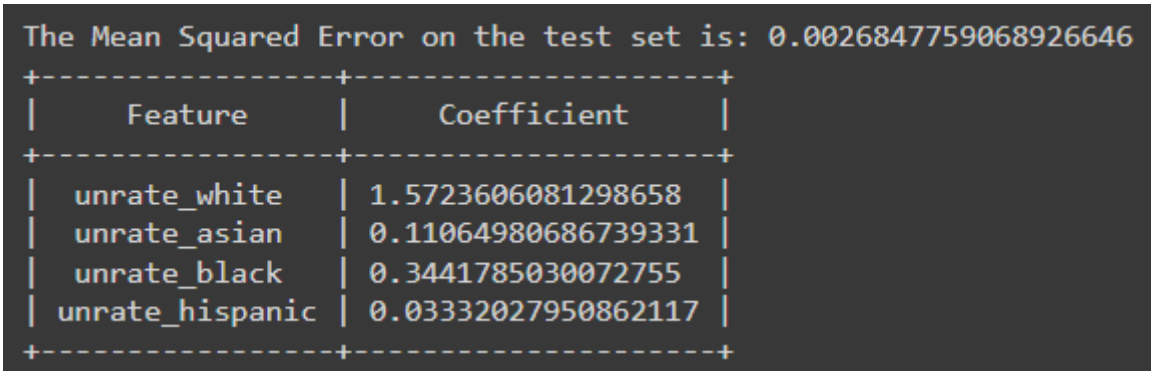
Multiple Linear Regression Prediction Equation for features with gender:
unrate = 0.00 + (0.47 * unrate_women) + (0.53 * unrate_men)

The multiple linear regression analysis conducted on the unemployment data reveals insightful relationships between different demographic groups' unemployment rates and the overall unemployment rate. The coefficients obtained indicate that increases in unemployment rates among white, black, and Asian individuals correspond to notable rises in the overall unemployment rate, while the impact of the Hispanic unemployment rate appears to be comparatively weaker. These findings are supported by the extremely low p-values, signifying the statistical significance of the relationships. Both 'unrate_women' and 'unrate_men' have statistically significant coefficients with extremely low p-values, indicating their significant impact on the overall unemployment rate. The positive coefficients suggest that increases in the unemployment rates among both men and women are associated with higher overall unemployment rates, with men having a slightly stronger impact compared to women.

4.3 Lasso Regression:

4.3.1 Lasso Regression Based on Races:

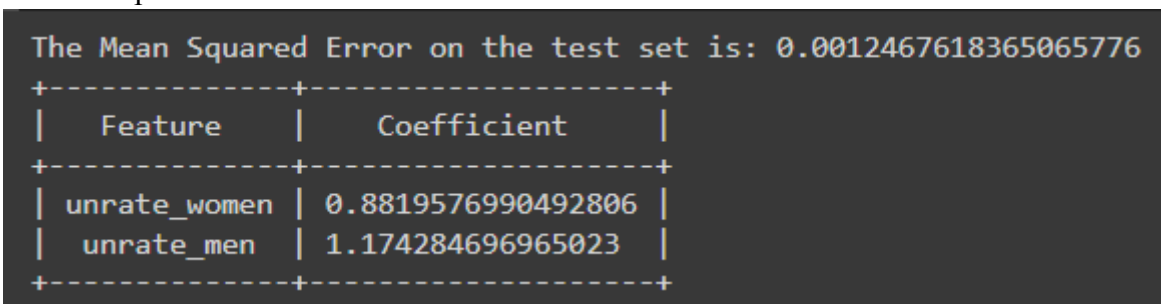
Table no. 4: Summary of Multiple Linear Regression Analyses performed for features with gender as input.



Lasso Regression Prediction Equation for features with race:
unrate = 5.97 + 1.57 * unrate_white + 0.11 * unrate_asian + 0.34 * unrate_black + 0.03 * unrate_hispanic

4.3.2 Lasso Regression Based on Gender:

Table no. 5: Summary of Multiple Linear Regression Analyses performed for features with gender as input.



Lasso Regression Prediction Equation for features with gender:
unrate = 5.97 + 0.88 * unrate_women + 1.17 * unrate_men

The image presents results from two Lasso Regression analyses that quantify the influence of race and gender on unemployment rates, separately. In the analysis focused on race, the coefficients suggest that the impact of unemployment rates varies significantly among racial groups, with whites showing the highest influence, followed by blacks, Asians, and Hispanics. The model achieves a low Mean Squared Error (MSE), indicating a good fit. Conversely, the gender-based analysis reveals that unemployment impacts men slightly more than women, also demonstrated by a very low MSE, suggesting the model's predictive accuracy is high. Both tables underscore the disparities in unemployment influences across different demographics, highlighting the varying economic experiences among these groups.

4.4 Support Vector Machine:

Table no.: 6 Summary of Multiple Linear Regression Analyses performed for features with races and gender as input.

Feature	Coefficient
unrate_white	0.5480568795759815
unrate_asian	0.0195624145399233
unrate_black	0.18993989901285213
unrate_hispanic	0.2115305311424771
unrate_women	0.5330333774308481
unrate_men	0.5653245886489562

The Mean Squared Error on the test set is: 0.001789599487730403

Support Vector Machine Prediction Equation for features:
0.548 * unrate_white + 0.020 * unrate_asian + 0.190 * unrate_black + 0.212 * unrate_hispanic + 0.533 * unrate_women + 0.565 * unrate_men + 5.964

Table no. 4 displays a table of coefficients from a regression analysis with the Mean Squared Error (MSE) of the model reported below the table. The coefficients represent the influence of unemployment rates across different demographic groups—white, Asian, black, Hispanic, women, and men—on a dependent variable. Notably, the coefficient for unemployment among white individuals is the highest at approximately 0.5409, suggesting a strong influence on the model's outcome, followed by men and women with coefficients of 0.5653 and 0.5330 respectively. The lowest impact is seen in the Asian group, with a coefficient of roughly 0.0196. The model achieves a very low MSE of 0.0018 on the test set, indicating high predictive accuracy.

This section summarizes the application of a Support Vector Machine (SVM) model to predict unemployment rates. The analysis focused on features including unemployment rates for different racial and gender groups.
The key findings are:
Feature Importance: The SVM model assigned the highest coefficients (positive influence on predicted unemployment) to unrate_white (0.548), unrate_women (0.533), and unrate_men (0.565). This suggests a stronger association between these features and the overall unemployment rate compared to unrate_asian (0.020) and unrate_hispanic (0.212).
Model Performance: The Mean Squared Error (MSE) on the test set was 0.0017, indicating a good fit between the model and unseen data.
Prediction Equation: The analysis produced an SVM prediction equation that can be used to estimate unemployment rates based on the values of the included features.

The SVM model successfully identified relationships between unemployment rates across racial and gender groups. The features with the highest coefficients (unrate_white, unrate_women, unrate_men) can be considered the most influential factors for predicting unemployment in this dataset. The low MSE suggests the model has the potential to make accurate predictions on new data. The provided equation allows for unemployment rate estimation based on the unemployment rates of the specific groups included in the model.

5.0 ANOVA

```
One-way ANOVA results:
F-statistic: 151.46
P-value: 0.0000
The one-way ANOVA test suggests that there is a significant difference between at least two groups.
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
unrate	unrate_asian	-1.1996	0.0	-1.8184	-0.5808	True
unrate	unrate_black	4.2344	0.0	3.6156	4.8531	True
unrate	unrate_hispanic	1.4668	0.0	0.848	2.0856	True
unrate	unrate_men	0.1917	0.9704	-0.4271	0.8105	False
unrate	unrate_white	-0.664	0.0261	-1.2828	-0.0453	True
unrate	unrate_women	-0.215	0.9482	-0.8338	0.4037	False
unrate_asian	unrate_black	5.434	0.0	4.8152	6.0528	True
unrate_asian	unrate_hispanic	2.6664	0.0	2.0476	3.2852	True
unrate_asian	unrate_men	1.3913	0.0	0.7725	2.0101	True
unrate_asian	unrate_white	0.5356	0.1408	-0.0832	1.1543	False
unrate_asian	unrate_women	0.9846	0.0001	0.3658	1.6033	True
unrate_black	unrate_hispanic	-2.7676	0.0	-3.3863	-2.1488	True
unrate_black	unrate_men	-4.0427	0.0	-4.6614	-3.4239	True
unrate_black	unrate_white	-4.8984	0.0	-5.5172	-4.2797	True
unrate_black	unrate_women	-4.4494	0.0	-5.0682	-3.8306	True
unrate_hispanic	unrate_men	-1.2751	0.0	-1.8939	-0.6563	True
unrate_hispanic	unrate_white	-2.1308	0.0	-2.7496	-1.5121	True
unrate_hispanic	unrate_women	-1.6818	0.0	-2.3006	-1.0631	True
unrate_men	unrate_white	-0.8557	0.0009	-1.4745	-0.237	True
unrate_men	unrate_women	-0.4067	0.4538	-1.0255	0.212	False
unrate_white	unrate_women	0.449	0.3284	-0.1697	1.0678	False

The table presents results from a Tukey HSD post-hoc test that we conducted to evaluate differences in unemployment rates across various demographic groups, following a significant finding from a one-way ANOVA. The test results highlight significant disparities in unemployment rates, with notable differences such as between Asians and blacks (mean difference 5.434, p = 0.0), Asians and Hispanics (mean difference 2.6664, p = 0.0), and blacks and whites (mean difference -4.8894, p = 0.0), all indicating statistically significant disparities. These results strongly reject the null hypothesis of equal unemployment rates among these groups, underscoring significant economic inequalities. In contrast, other group comparisons like Asians vs. whites (mean difference 0.5356, p = 0.1408) and men vs. women (mean difference -0.4067, p = 0.4538) did not reveal statistically significant differences, suggesting similar unemployment rates between these demographics. The "reject" column clearly indicates which comparisons have significant disparities, with many entries marked true, confirming the presence of significant differences. This analysis provides a comprehensive view of employment challenges across demographic groups, revealing both significant disparities and similarities where the unemployment rates appear comparably stable.

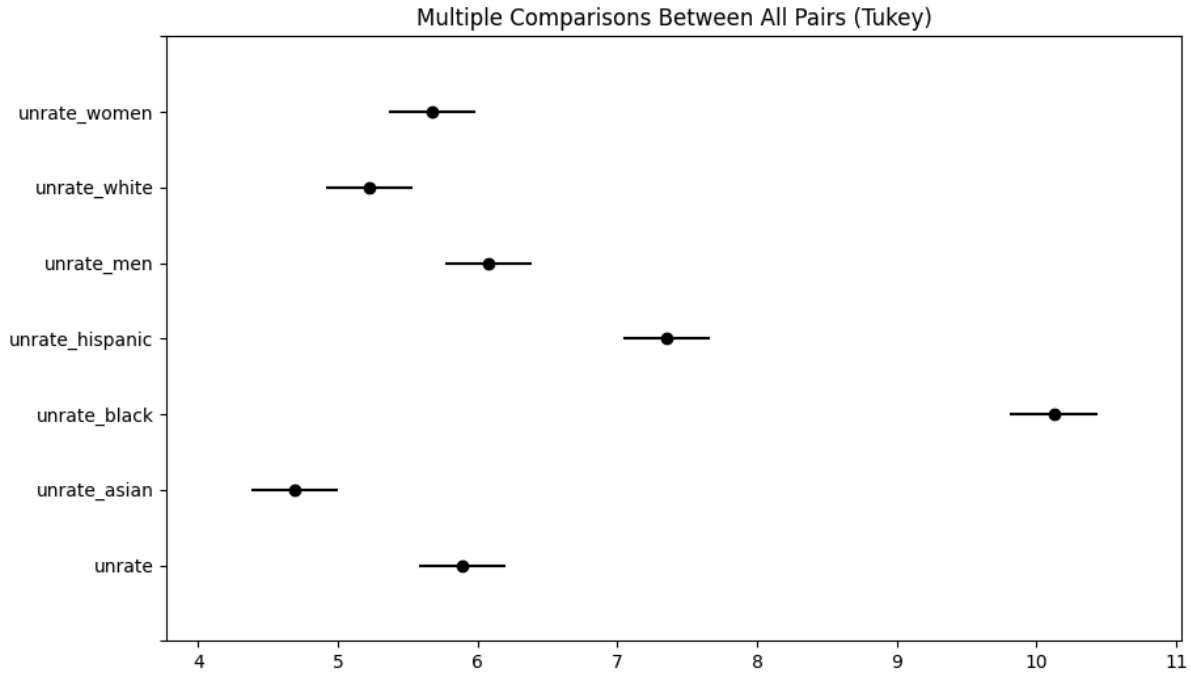


Figure 12. The Tukey Honestly Significant Difference (HSD) test detects variance among group means.

The image shows a multiple comparisons plot from a Tukey HSD test, illustrating confidence intervals for the means of unemployment rates across various demographic groups. Each horizontal line represents a confidence interval for a group, with the dot denoting the group's mean unemployment rate. The groups displayed are general, Asian, black, Hispanic, men, white, and women, listed from bottom to top. The Asian group has the lowest mean unemployment rate, indicated by its position near the bottom of the plot and centered around 5. The black group has the highest mean, centered around 9.5, showing a clear disparity in unemployment rates. Groups like women and men are close to each other, centered around 6.5 and 6.7 respectively, suggesting similar unemployment rates. This visual representation effectively highlights the differences and similarities in unemployment rates between these groups, making it easier to identify which groups are more affected.

6.0 Summary and Conclusion

The analysis conducted on unemployment rates in the USA revealed several significant insights into the dynamics of unemployment across different demographic groups. Initially, we explored the overall unemployment trend, identifying two major spikes corresponding to the Great Recession in 2009 and the COVID-19 pandemic in 2020. These spikes highlighted the economy's vulnerability to external shocks and the subsequent recovery efforts.

Subsequently, we delved into unemployment patterns among various ethnic and gender groups. We observed distinct trends, such as the disproportionate impact of economic downturns on different genders and ethnicities. For instance, during the Great Recession, men experienced higher unemployment rates due to their concentration in industries affected by the housing market collapse. Conversely, the COVID-19 crisis led to a higher job loss among

women, mainly attributed to their representation in lower-paying sectors susceptible to pandemic-related disruptions.

In the inferential analysis, we investigated the correlation between demographic factors and unemployment rates. Simple and multiple linear regression models were employed to quantify the relationships between unemployment rates among different demographic groups and the overall unemployment rate. The results highlighted significant correlations between unemployment rates among white, black, and Asian individuals and the overall unemployment rate, with women and men also contributing significantly.

In conclusion, our study underscores the importance of considering demographic factors such as gender and ethnicity in understanding unemployment dynamics. By developing regression models and analyzing historical trends, we can gain valuable insights into the factors driving unemployment and make informed decisions to mitigate its impact. Moving forward, policymakers and stakeholders can leverage these insights to implement targeted interventions and policies aimed at reducing unemployment disparities and fostering inclusive economic growth.

7.0 Future Work

While our study provides valuable insights into unemployment dynamics and its correlation with demographic factors, there are several avenues for future research to explore further. Some potential directions for future work include:

1. Geospatial Analysis: Integrating geospatial analysis to examine regional variations in unemployment rates and identify localized factors contributing to disparities in employment outcomes. Geospatial techniques can help policymakers target interventions and resources more effectively in areas with the highest unemployment rates and socioeconomic challenges. By addressing these aspects in future research, we can enhance our understanding of the drivers of unemployment and develop more accurate and actionable predictive models to support evidence-based policymaking and economic planning efforts.

2. Causal Inference Analysis: Causal inference analysis aims to identify causal relationships between specific interventions, policies, or economic factors and changes in unemployment rates. Techniques such as propensity score matching, difference-in-differences, and instrumental variable analysis can help evaluate the causal impact of policy interventions, economic shocks, or structural changes on unemployment outcomes.