Interaction Effects on Cancer Death Rates

Introduction

This report examines the interaction effects between various socioeconomic and demographic factors on cancer death rates across different regions. Specifically, we analyze how interactions between variables such as Socioeconomic Status (SES), education, unemployment, race, and urbanicity influence the relationship between poverty and cancer death rates. By exploring these interaction effects, we aim to uncover more complex relationships that may not be evident when examining individual variables independently.

Methodology

Data Preparation

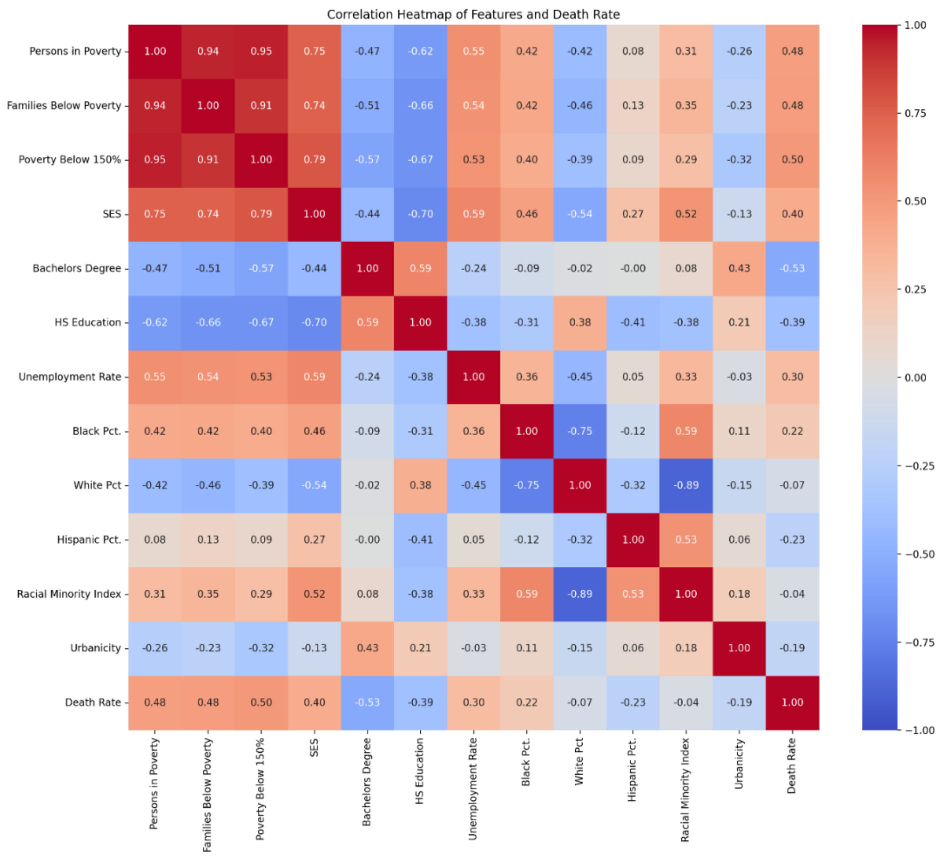
The initial dataset included the following variables: Persons in Poverty, Families Below Poverty, Poverty Below 150%, SES, Bachelor's Degree, High School Education, Unemployment Rate, Black Percentage, White Percentage, Hispanic Percentage, Racial Minority Index, and Urbanicity. Before running the regression analysis, we performed data cleaning to address any non-numeric values and standardized categorical variables.

Addressing Multicollinearity

Multicollinearity occurs when independent variables in a regression model are highly correlated, making it difficult to interpret the individual effects of each variable. To address this issue, we used the following steps:

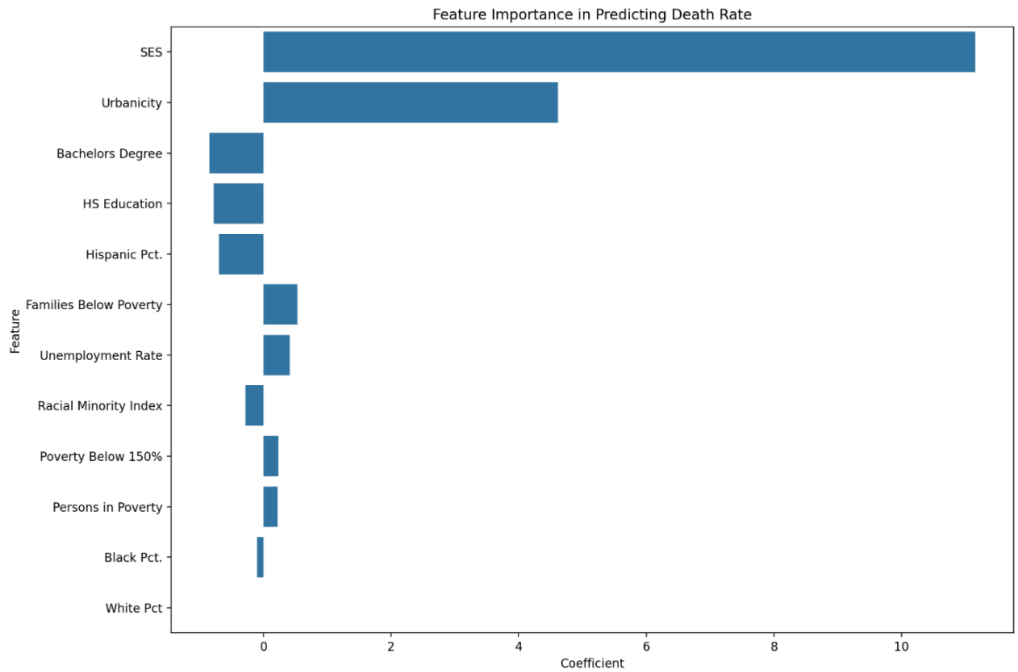
1. **Correlation Analysis:** We generated a correlation matrix to identify pairs of variables that were highly correlated (e.g., SES and Poverty Below 150%).

Figure 1: Correlation Matrix of Initial Features



2. **Variance Inflation Factor (VIF):** We calculated VIF values for all variables to quantify the severity of multicollinearity. VIF values above 5-10 indicate problematic multicollinearity.

Figure 2: Initial Variance Inflation Factors (VIF)



3. **Feature Selection:** Based on the correlation analysis and VIF values, we removed highly correlated variables to reduce multicollinearity. The final set of variables included SES, Bachelor's Degree, High School Education, Unemployment Rate, Black Percentage, Hispanic Percentage, and Urbanicity.

Figure 3: Updated Correlation Matrix of Selected Features

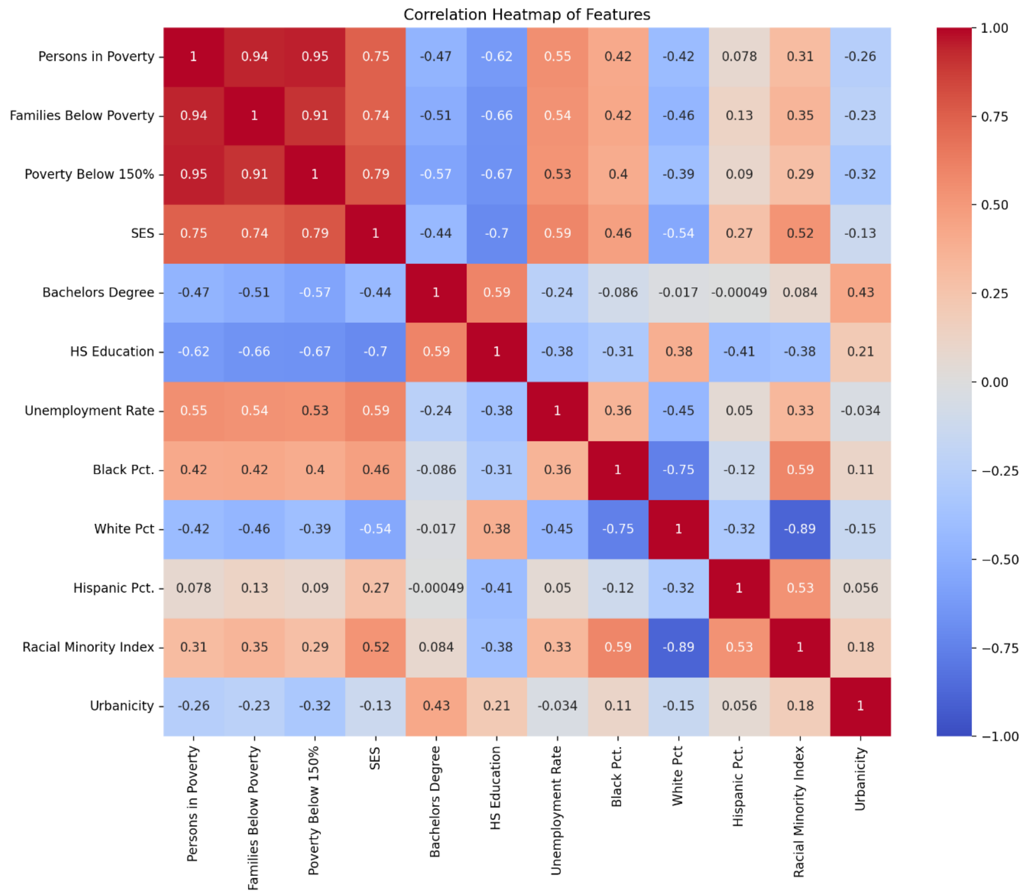
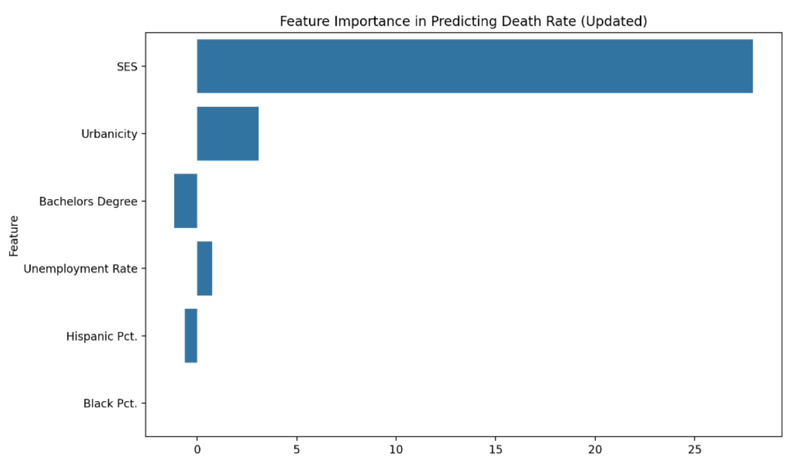


Figure 4: Updated Variance Inflation Factors (VIF)



Results

With the reduced set of features, we conducted a multiple regression analysis with cancer death rate as the dependent variable. The results are as follows:

Model Performance

Mean Squared Error: 418.4246

R-squared: 0.4340

The R-squared value of 0.4340 indicates that approximately 43.40% of the variance in cancer death rates can be explained by the selected features. Although the R-squared value decreased slightly after reducing multicollinearity, the model's interpretability has improved.

Feature Coefficients

Figure 4: Feature Coefficients in the Multiple Regression Model

Feature	Coefficient
SES	27.93
Urbanicity	3.08
Bachelor's Degree	-1.14
Unemployment Rate	0.76
Hispanic Pct.	-0.62
Black Pct.	-0.01

Interpretation of Results

Main Effects

SES: SES remains a significant predictor of cancer death rates, with a strong positive relationship. This suggests that areas with higher socioeconomic status tend to have higher death rates, which is counterintuitive and requires further investigation.

Urbanicity: Urban areas tend to have higher cancer death rates, due to factors such as environmental pollution, stress, and lifestyle differences between urban and rural areas.

Education and Race: Higher education levels (Bachelor's Degree) and higher Hispanic populations are associated with lower death rates, which aligns with the "Hispanic Paradox" observed in public health literature.

Exploring Interaction Effects

To further understand the complex relationships between these variables, we explored interaction effects between SES and other key predictors (Bachelor's Degree, Unemployment Rate, and Urbanicity). Interaction terms help us understand how the relationship between SES and cancer death rates varies across distinct levels of these variables.

Model Performance with Interaction Effects

Mean Squared Error: 405.5724

R-squared: 0.4514

Including interaction terms improved the model's explanatory power, as indicated by an increase in the R-squared value from 0.4340 to 0.4514.

Interaction Effects

Figure 5: Feature Coefficients (Including Interaction Terms)

Feature	Coefficient
SES	13.49
Bachelor's Degree	-10.26
Hispanic Pct.	-8.58
Urbanicity	4.81
SES x Urbanicity	-3.88
Unemployment Rate	3.48
SES x Unemployment	-3.35
SES x Bachelor's Degree	-2.59
Black Pct.	0.05

Figure 6: Feature Importance in Predicting Death

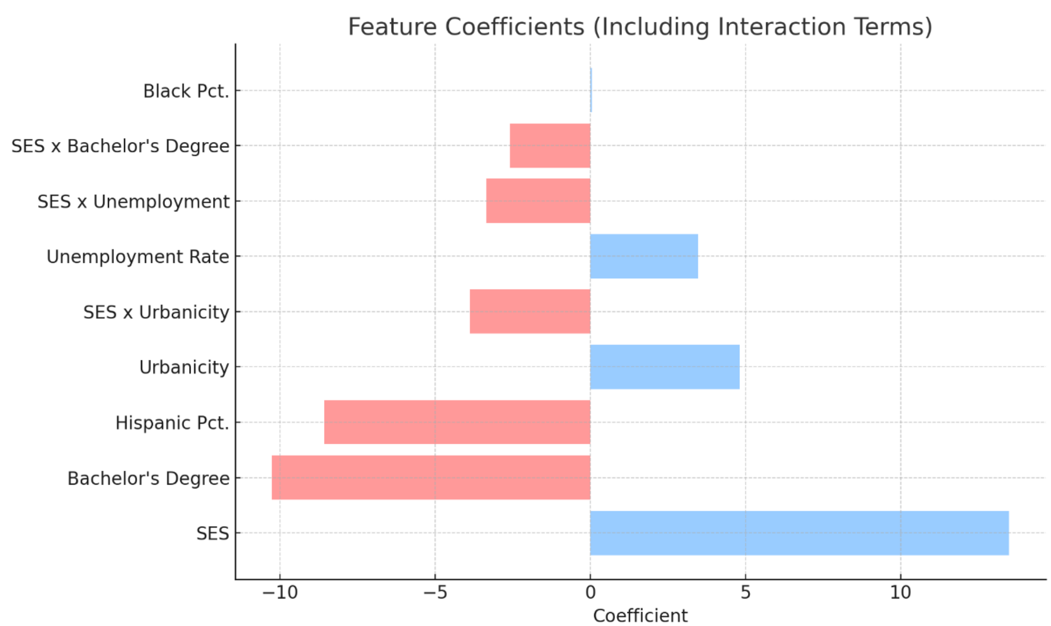
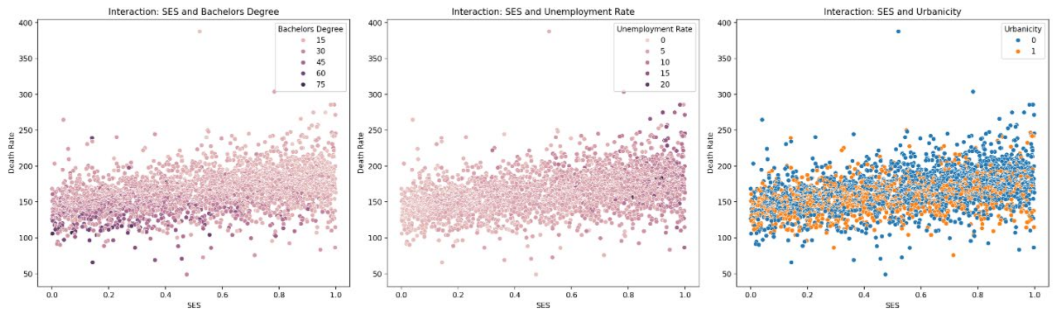


Figure7: Interaction Effects Visualization



Interpretation of Interaction Effects

1. **SES x Urbanicity:** The positive effect of SES on cancer death rates is less pronounced in urban areas. This suggests that urban areas may have resources or characteristics that mitigate some of the negative impacts associated with higher SES.

2. **SES x Unemployment:** The effect of SES on cancer death rates is reduced in areas with higher unemployment. This indicates that unemployment may have a more uniform effect across different SES levels.

3. **SES x Bachelor's Degree:** Higher education levels reduce the positive relationship between SES and cancer death rates, implying that education has a protective effect, particularly in areas with higher SES.

Implications and Next Steps

The findings from this analysis reveal that the relationship between SES and cancer death rates is not uniform and is influenced by other factors such as education, unemployment, and urbanicity. These results have several important implications:

Targeted Interventions: Public health interventions should be tailored to account for the complex interactions between SES, education, and urbanicity. For example, strategies to reduce cancer death rates in urban areas may differ from those in rural areas due to the different dynamics at play.

Education's Protective Effect: The strong negative interaction between SES and Bachelor's Degree highlights the importance of education in improving health outcomes, suggesting that policies promoting higher education levels could be effective in reducing cancer death rates.

Further Investigation: The counterintuitive positive relationship between SES and cancer death rates, even after accounting for interactions, warrants further investigation to identify potential confounding factors or measurement issues.

Conclusion

This analysis has provided a deeper understanding of how interaction effects between socioeconomic and demographic variables influence cancer death rates. By accounting for these interactions, we have gained valuable insights that can inform more effective public health strategies. Future research should continue to explore these complex relationships and consider additional variables that may further elucidate the determinants of cancer outcomes.