

## **Project Proposal: Identifying Key Socioeconomic and Demographic Factors for Predictive Health Equity Modeling in Medically Underserved Areas**

### **Objective:**

Our project aims to understand the socioeconomic factors associated with the designation of Medically Underserved Areas/Populations (MUA/P) at County level in the United States. By analyzing how these factors contribute and influence MUA/P status, we seek to identify key predictors of underserved status, develop a model to predict MUA/P designation, and uncover clusters of regions with similar healthcare needs. Ultimately, we hope this analysis will inform efforts to address healthcare access disparities in underserved regions.

### **Research Questions and Methods:**

1. **Q1: What are the geographic distribution patterns of MUA/P in the U.S., and which regions are most affected?**

**Method:** We will use geocoding to visualize the distribution pattern of Medically Underserved and well-served areas across the United States, both at State and County levels.

2. **Q2: Which socioeconomic factors are most strongly linked to MUA/P designation?**

**Method:** We will build a logistic regression model with stepwise forward feature selection based on AIC to determine what socioeconomic factors are most influential to MUA/P designation among the 20ish features in the dataset such as poverty levels, unemployment rates, and education levels, to prevent multicollinearity and overfitting.

3. **Q3: Can we predict MUA/P designations using selected features?**

**Method:** We will train a Random Forest classifier on 2021-2022 data, test on 2023 data, and predict MUA/P status for 2024. Evaluate model accuracy and analyze feature importance.

4. **Q4: What factors contribute to changes in MUA/P status over time?**

**Method:** We will apply a Difference-in-Differences approach to compare changes in key variables for counties experiencing shifts in MUA/P status versus stable counties. This will analyze trends and the impact of variables over three years.

### **Data Sources and Key Variables:**

1. **Medically Underserved Area/Population (MUA/P)**
  - **Source:** HRSA's MUA/P Designation Database
  - **Key Variables:** Population data, provider ratios, income levels, age, geo-codes, designation dates.
2. **Kaiser Family Foundation (KFF) State Data**
  - **Source:** KFF State Health Facts
  - **Key Variables:** Population size, race/ethnicity distribution, poverty levels, median income, unemployment, health insurance coverage.
3. **American Community Survey (ACS)**

- **Source:** U.S. Census Bureau
- **Key Variables:** Educational attainment levels.

**Code:**

```
mua=pd.read_csv('MUA_DET.csv')|mua_county=mua.groupby(['County','State']).sum(numeric_only=True)
```

```
acs=pd.read_csv('ACS.csv',header =1| est_col = list(filter(lambda x: 'Estimate' in x, a)) |
est_tot_col = list(filter(lambda x: 'Total' in x, est_col))| edu_col = list(filter(lambda x:
'EDUCATION' in x, est_tot_col))|
```

```
race_col = list(filter(lambda x: 'One race' in x, est_tot_col))| house_col = list(filter(lambda x:
'HOUSING TENURE' in x, est_tot_col))|pov_col = list(filter(lambda x: 'POVERTY STATUS' in x,
est_tot_col))| needed_col = ['Geographic Area
Name']+edu_col+race_col+house_col+pov_col|acs = acs[needed_col]
```

County	State	edu_Total_Pop_25+	edu_Less_HighSchool	edu_HighSchool_Graduate	edu_Some_College_Associate	edu_Bachelors_Degree	edu_Graduate_Professional_Degree
Ada County	Idaho	364782.0	16853.0	75079.0	111707.0	104908.0	56235.0
Adams County	Colorado	355595.0	46857.0	103406.0	99248.0	71793.0	34291.0
	Pennsylvania	75880.0	6362.0	30487.0	20916.0	10803.0	7312.0
Aiken County	South Carolina	124842.0	11409.0	38801.0	32802.0	27022.0	14808.0
Alachua County	Florida	174855.0	8185.0	31024.0	47492.0	37618.0	50536.0

	County	State	MUAP_Code	MUAP_Service_Area	MUAP_Status	Designation_Date	MUAP_Update_Date	MUAP_Withdrawal_Date	IMU_Score	Popula
0	Bexar County	Texas	48029180101	Bexar-San Antonio West	Designated	11/19/19	11/19/19	NaN	57.0	Ur
1	Bexar County	Texas	48029180604	Bexar-San Antonio West	Designated	11/19/19	11/19/19	NaN	57.0	Ur
2	Bexar County	Texas	48029180800	Bexar-San Antonio West	Designated	11/19/19	11/19/19	NaN	57.0	Ur
3	Bexar County	Texas	48029190501	Bexar-San Antonio West	Designated	11/19/19	11/19/19	NaN	57.0	Ur