SPRING 2025

Random Forest Analysis of Health and Poverty Dynamics in the US (2015-2025)

By
Juana Wong

CSc 59866 L
Senior Project I - Modeling of Complex Systems

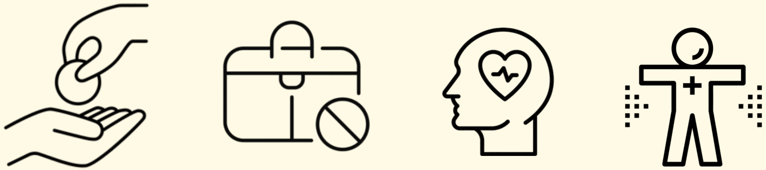
AGENDA

01

PROBLEM STATEMENT

This project analyzes county-level health and socioeconomic data across the United States from 2015–2025.

It examines trends in poverty, unemployment, and mental and physical health to identify key areas for improving quality of life nationwide.



DATASETS / DATA SOURCES

HEALTH, POVERTY, UNEMPLOYMENT, LOCATION

Population Dynamics Foundation Model (PDFM) Embeddings

- conus27.csv
- county_unemployment.csv
- zcta_poverty.csv

Google Research

11.9k followers · Earth · https://research.google

github.com/google-research

HEALTH, UNEMPLOYMENT, DEMOGRAPHICS, LOCATION

County-Level Population Health and Well-being and Community Conditions

2015 CHR CSV Analytic Data	2020 CHR Analytic Data (CSV)
2016 CHR CSV Analytic Data	2021 CHR CSV Analytic Data
2017 CHR CSV Analytic Data	2022 CHR CSV Analytic Data
2018 CHR CSV Analytic Data	2023 CHR CSV Analytic Data
2019 CHR CSV Analytic Data	2024 CHR CSV Analytic Data
	2025 CHR CSV Analytic Data

County Health Rankings & Roadmaps

ZIP CODES, FIPS CODES, STATES

FIPS codes for all U.S. locations in a spreadsheet

ZIP codes to FIPS codes mapping

The ZIP codes to FIPS codes sheet provides the list of USPS ZIP codes with their population and maps each ZIP code to state FIPS code and county FIPS code as well as CBSA codes for metropolitan statistical areas. You can use this sheet to lookup FIPS codes by ZIP code or conversely filter to lookup ZIP codes by FIPS code. Note that this sheet includes the primary county for each ZIP code. The ZIP code to county mapping below includes all unique ZIP to county relationships (some ZIP codes overlap with multiple counties).

File Edit View Data Insert Help — FIPS code lists and location mapping — Share

A1 — ZIP Code

	ZIP Code	State FIPS code	Primary County FIPS	Metro CBSA	Town	State Abbr	State	Primary County	Metro	Total population
2	35004	01	01115	13820	Moody	AL	Alabama	St. Clair County	Birmingham, AL	11,558
3	35005	01	01073	13820	Adamsville	AL	Alabama	Jefferson County	Birmingham, AL	8,143
4	35006	01	01073	13820	Adger	AL	Alabama	Jefferson County	Birmingham, AL	3,599
5	35007	01	01117	13820	Alabaster	AL	Alabama	Shelby County	Birmingham, AL	27,935
6	35010	01	01123	10760	Alexander City	AL	Alabama	Tallapoosa County	Alexander City, AL	19,619
7	35013	01	01009	13820	Allgood	AL	Alabama	Blount County	Birmingham, AL	122
8	35014	01	01121	45180	Alpine	AL	Alabama	Talladega County	Talladega-Sylacauga, AL	3,353
9	35016	01	01095	10700	Arab	AL	Alabama	Marshall County	Albertville, AL	17,435
10	35019	01	01043	18980	Baileyton	AL	Alabama	Cullman County	Cullman, AL	2,497
11	35020	01	01073	13820	Bessemer	AL	Alabama	Jefferson County	Birmingham, AL	25,308
12	35022	01	01073	13820	Bessemer	AL	Alabama	Jefferson County	Birmingham, AL	23,989
13	35023	01	01073	13820	Bessemer	AL	Alabama	Jefferson County	Birmingham, AL	25,221
14	35031	01	01009	13820	Blountsville	AL	Alabama	Blount County	Birmingham, AL	7,937
15	35032	01	01121	45180	Bon Air	AL	Alabama	Talladega County	Talladega-Sylacauga, AL	78

County FIPS codes · State FIPS codes · ZIP codes to FIPS codes · FIPS codes for U.S. places (curated) · Census places only

row zero

FIPS CODES, STATES

USPS State Abbreviations and FIPS Codes

State	Postal Abbr.	FIPS Code	State	Postal Abbr.	FIPS Code
Alabama	AL	01	Nebraska	NE	31
Alaska	AK	02	Nevada	NV	32
Arizona	AZ	04	New Hampshire	NH	33
Arkansas	AR	05	New Jersey	NJ	34
California	CA	06	New Mexico	NM	35
Colorado	CO	08	New York	NY	36
Connecticut	CT	09	North Carolina	NC	37
Delaware	DE	10	North Dakota	ND	38
District of Columbia	DC	11	Ohio	OH	39
Florida	FL	12	Oklahoma	OK	40
Georgia	GA	13	Oregon	OR	41
Hawaii	HI	15	Pennsylvania	PA	42
Idaho	ID	16	Puerto Rico	PR	72
Illinois	IL	17	Rhode Island	RI	44
Indiana	IN	18	South Carolina	SC	45
Iowa	IA	19	South Dakota	SD	46
Kansas	KS	20	Tennessee	TN	47
Kentucky	KY	21	Texas	TX	48
Louisiana	LA	22	Utah	UT	49
Maine	ME	23	Vermont	VT	50
Maryland	MD	24	Virginia	VA	51
Massachusetts	MA	25	Virgin Islands	VI	78
Michigan	MI	26	Washington	WA	53
Minnesota	MN	27	West Virginia	WV	54
Mississippi	MS	28	Wisconsin	WI	55
Missouri	MO	29	Wyoming	WY	56
Montana	MT	30			

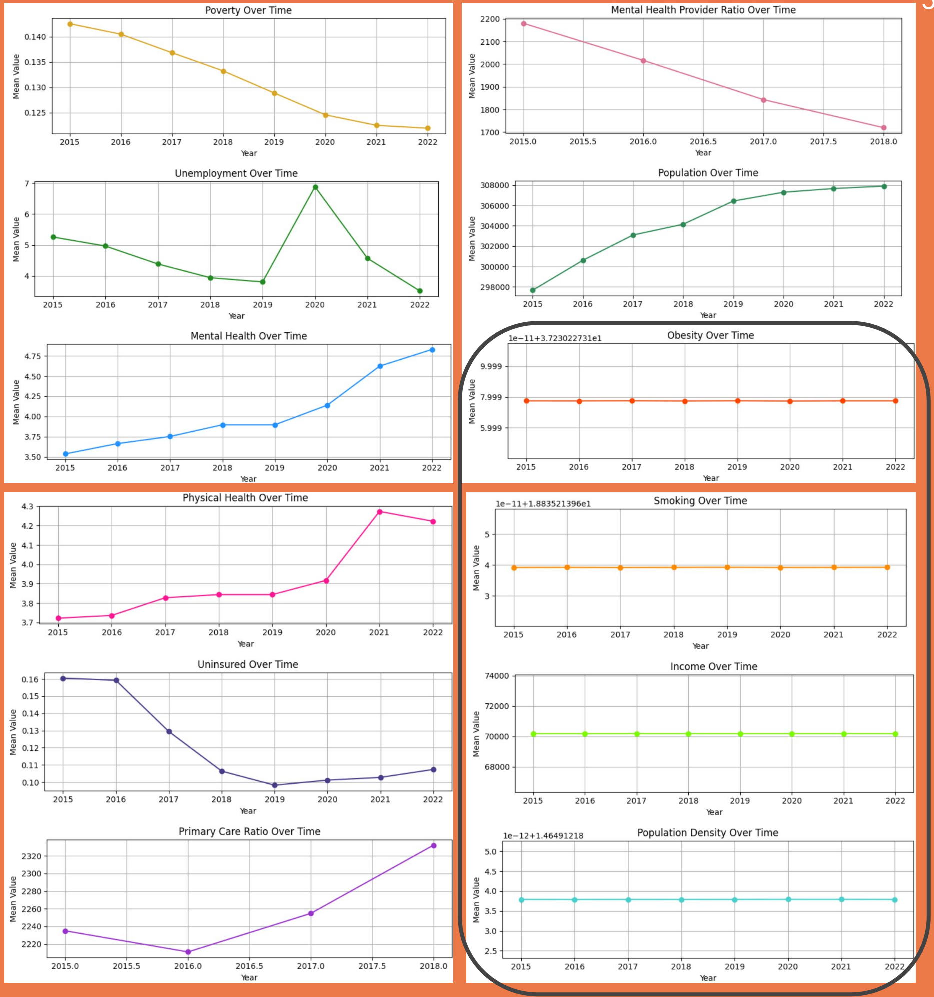
U.S. BUREAU OF LABOR STATISTICS

EXPLORATORY DATA ANALYSIS (EDA)

```
[5 rows x 32 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35126 entries, 0 to 35125
Data columns (total 32 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   fips5                          35126 non-null  object
 1   year                           35126 non-null  int64
 2   poor_mental_health_days        24983 non-null  float64
 3   poor_physical_health_days      25201 non-null  float64
 4   uninsured                      35107 non-null  float64
 5   ratio_primary_care_physicians  12222 non-null  float64
 6   ratio_mental_health_providers  11568 non-null  float64
 7   unemployment                   35105 non-null  float64
 8   population                     35117 non-null  float64
 9   suicides                       14822 non-null  float64
 10  crude_suicide_rate             14822 non-null  float64
 11  frequent_mental_distress       9580 non-null   float64
 12  frequent_physical_distress     9580 non-null   float64
 13  poor_or_fair_health            9580 non-null   float64
 14  life_expectancy                9363 non-null   float64
 15  diabetes_prevalence            9580 non-null   float64
 16  hiv_prevalence                 8208 non-null   float64
 17  drug_overdose_deaths           5859 non-null   float64
 18  insufficient_sleep             9577 non-null   float64
 19  adult_smoking                  9580 non-null   float64
 20  adult_obesity                  9580 non-null   float64
 21  physical_inactivity            9580 non-null   float64
 22  excessive_drinking             9580 non-null   float64
 23  preventable_hospital_stays     9366 non-null   float64
 24  children_in_poverty            9579 non-null   float64
 25  median_household_income        9578 non-null   float64
 26  income_inequality              9538 non-null   float64
 27  air_pollution_pm               6332 non-null   float64
 28  drinking_water_violations      9431 non-null   float64
 29  traffic_volume                 9366 non-null   float64
 30  pct_below_18                   9582 non-null   float64
 31  pct_65_and_older               9582 non-null   float64
dtypes: float64(30), int64(1), object(1)
memory usage: 8.6+ MB
```

```
--------------MISSING--------------
fips5                              0
year                               0
poor_mental_health_days        10143
poor_physical_health_days       9925
uninsured                         19
ratio_primary_care_physicians  22904
ratio_mental_health_providers  23558
unemployment                      21
population                         9
suicides                       20304
crude_suicide_rate             20304
frequent_mental_distress       25546
frequent_physical_distress     25546
poor_or_fair_health            25546
life_expectancy                25763
diabetes_prevalence            25546
hiv_prevalence                 26918
drug_overdose_deaths           29267
insufficient_sleep             25549
adult_smoking                  25546
adult_obesity                  25546
physical_inactivity            25546
excessive_drinking             25546
preventable_hospital_stays     25760
children_in_poverty            25547
median_household_income        25548
income_inequality              25588
air_pollution_pm               28794
drinking_water_violations      25695
traffic_volume                 25760
pct_below_18                   25544
pct_65_and_older               25544
dtype: int64
```

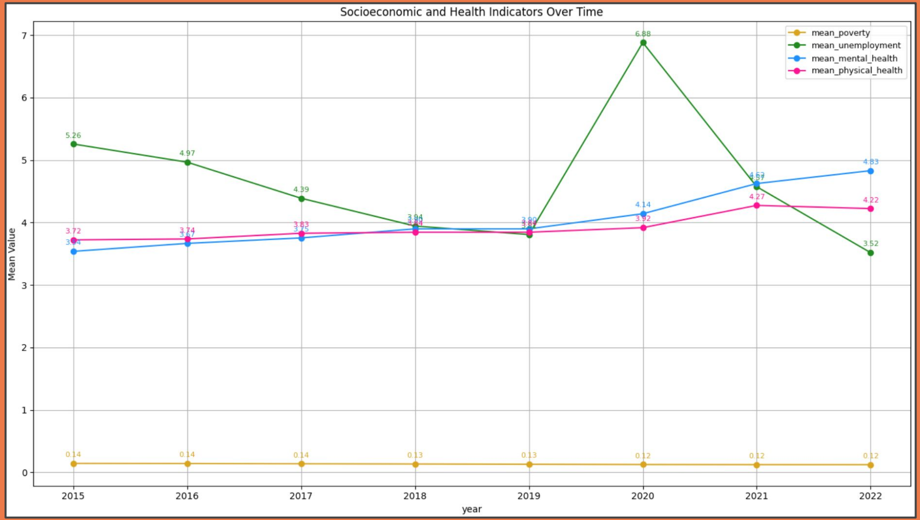
County Health Rankings

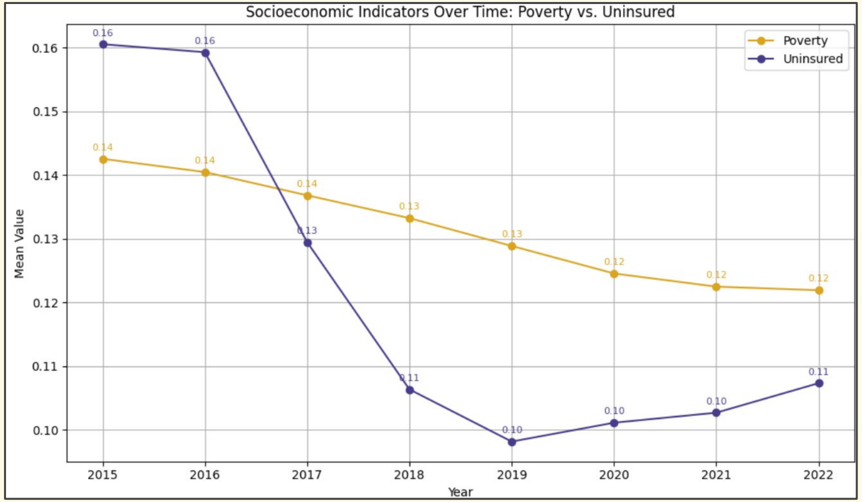


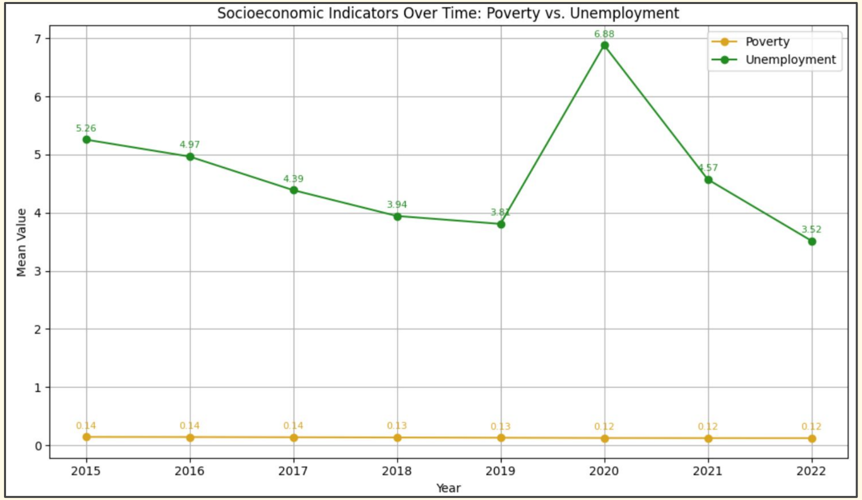
conus27

03

TIME SERIES (EDA)



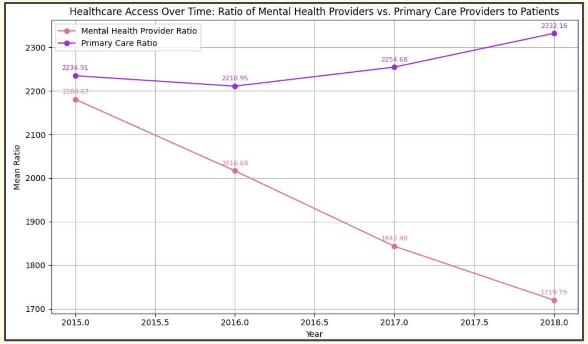


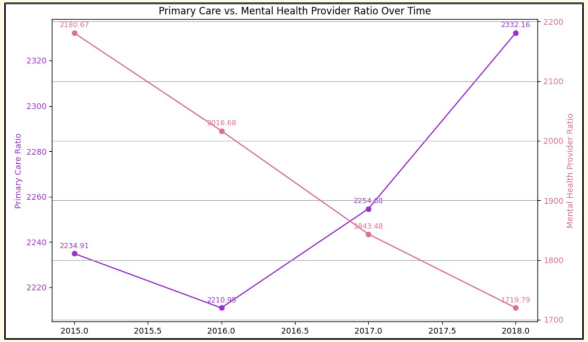


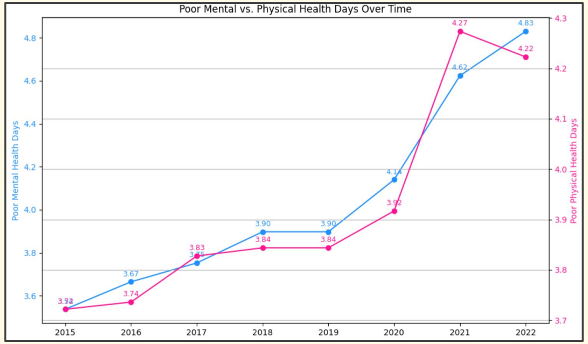
- **Economic variables** (poverty, unemployment, uninsured) are largely influenced by other unknown factors

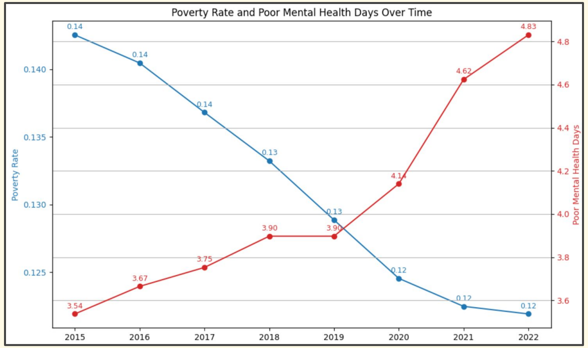
03

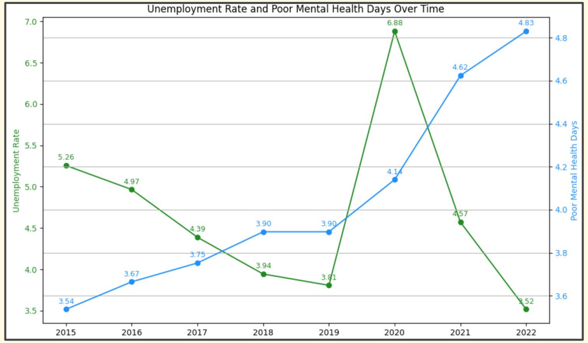
TIME SERIES (EDA)

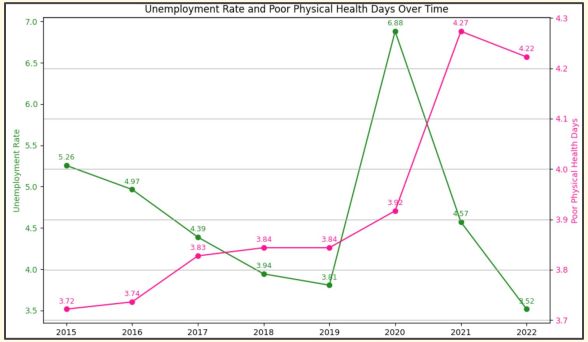








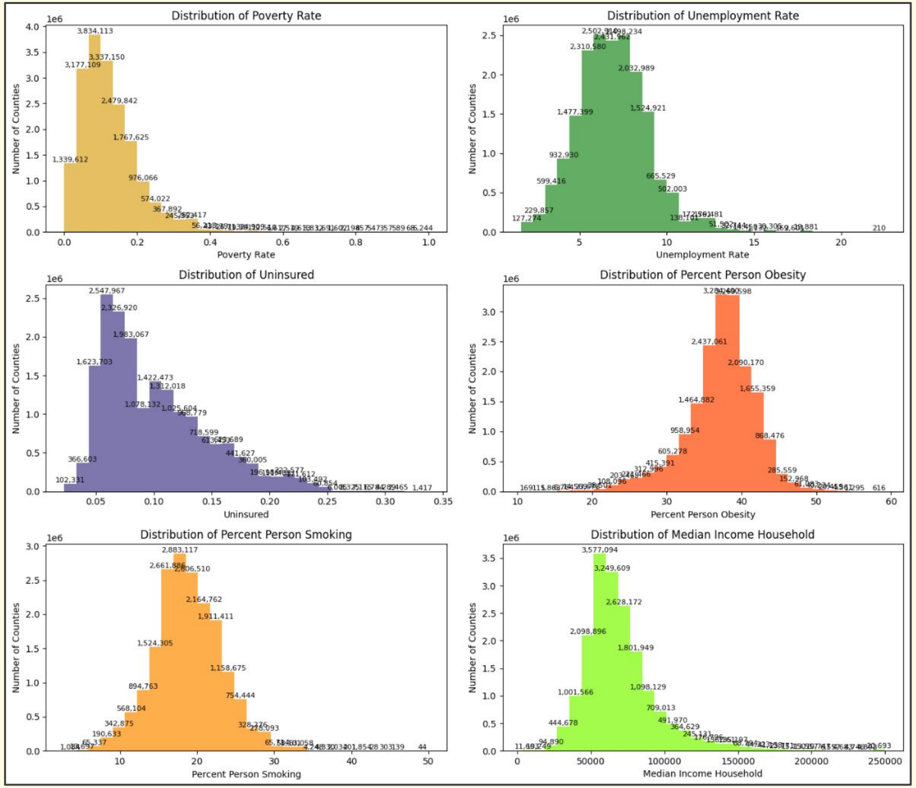


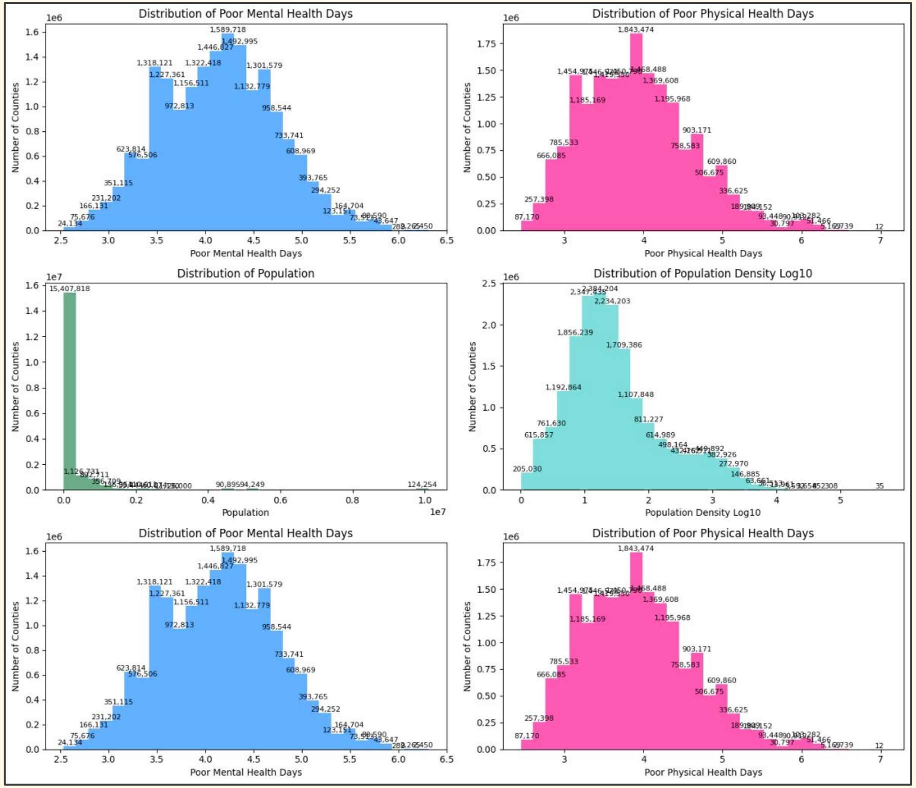


- **Mental** and **physical health** appear closely related.
- **Access to primary (health) care** *increases* as **access to mental healthcare** *decreases* from 2015 to 2018
(no data available for 2019-2025).
- **Economic factors** do not appear to have significant effect on **poor mental** or **physical health days**.

FEATURE DISTRIBUTION

by num counties, for all years





**

STRONG POSITIVE RELATIONSHIPS, >= +0.50

poor_mental_health_days & **poor_physical_health_days: 0.94**

Percent_Person_Obesity & **Percent_Person_Smoking: 0.75**

poor_physical_health_days & **Percent_Person_Smoking: 0.70**

poor_mental_health_days & **Percent_Person_Smoking: 0.66**

MODERATE POSITIVE RELATIONSHIPS, >= 0

poverty_rate & poor_physical_health_days: 0.50

poverty_rate & poor_mental_health_days: 0.44

poverty_rate & Percent_Person_Smoking: 0.42

unemployment_rate & poor_mental_health_days: 0.40

poor_physical_health_days & Percent_Person_Obesity: 0.45

poor_mental_health_days & Percent_Person_Obesity: 0.39

MODERATE NEGATIVE RELATIONSHIPS, <= 0

poor_physical_health_days & Median_Income_Household: -0.48

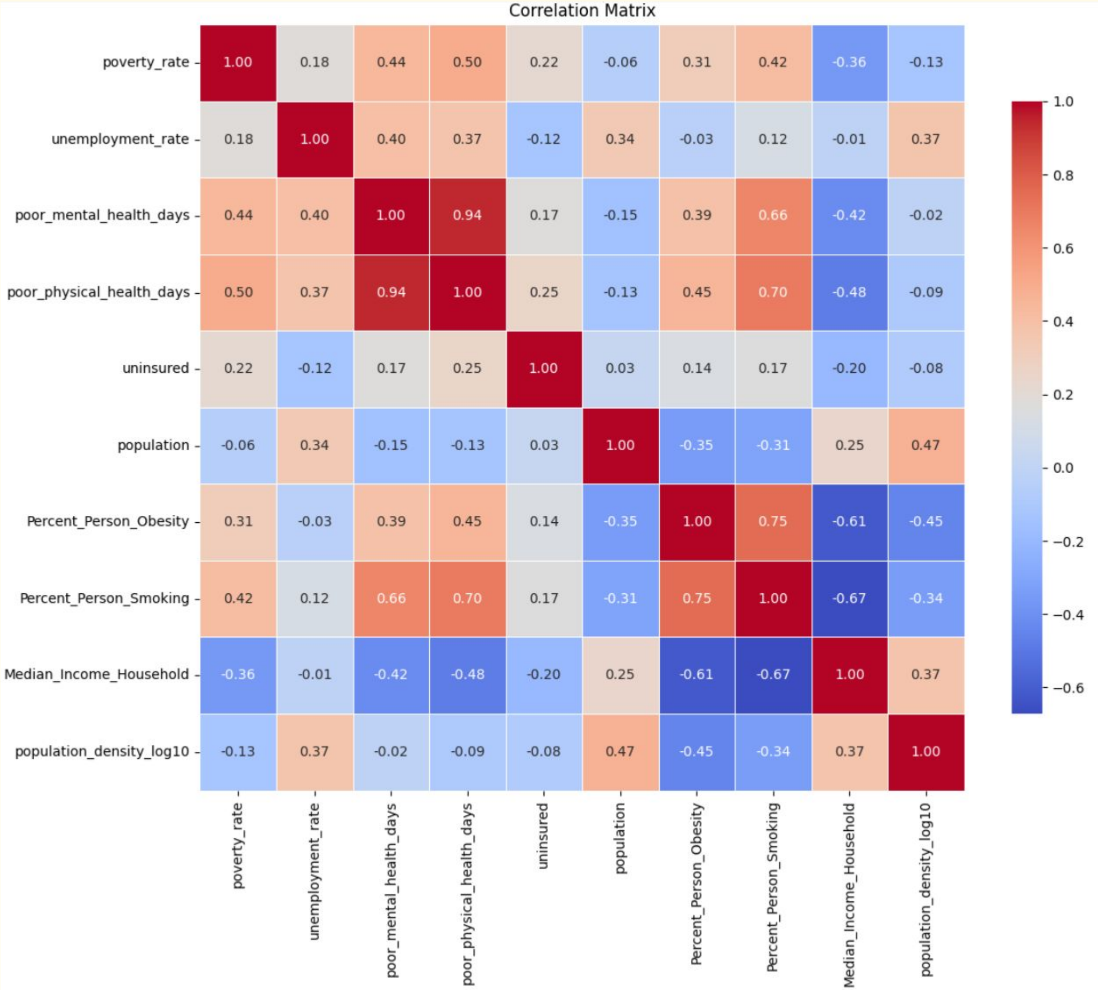
Percent_Person_Obesity & population_density_log10: -0.45

poor_mental_health_days & Median_Income_Household: -0.42

STRONG NEGATIVE RELATIONSHIPS, <= −0.50

Percent_Person_Smoking & Median_Income_Household: -0.67

Percent_Person_Obesity & Median_Income_Household: -0.61



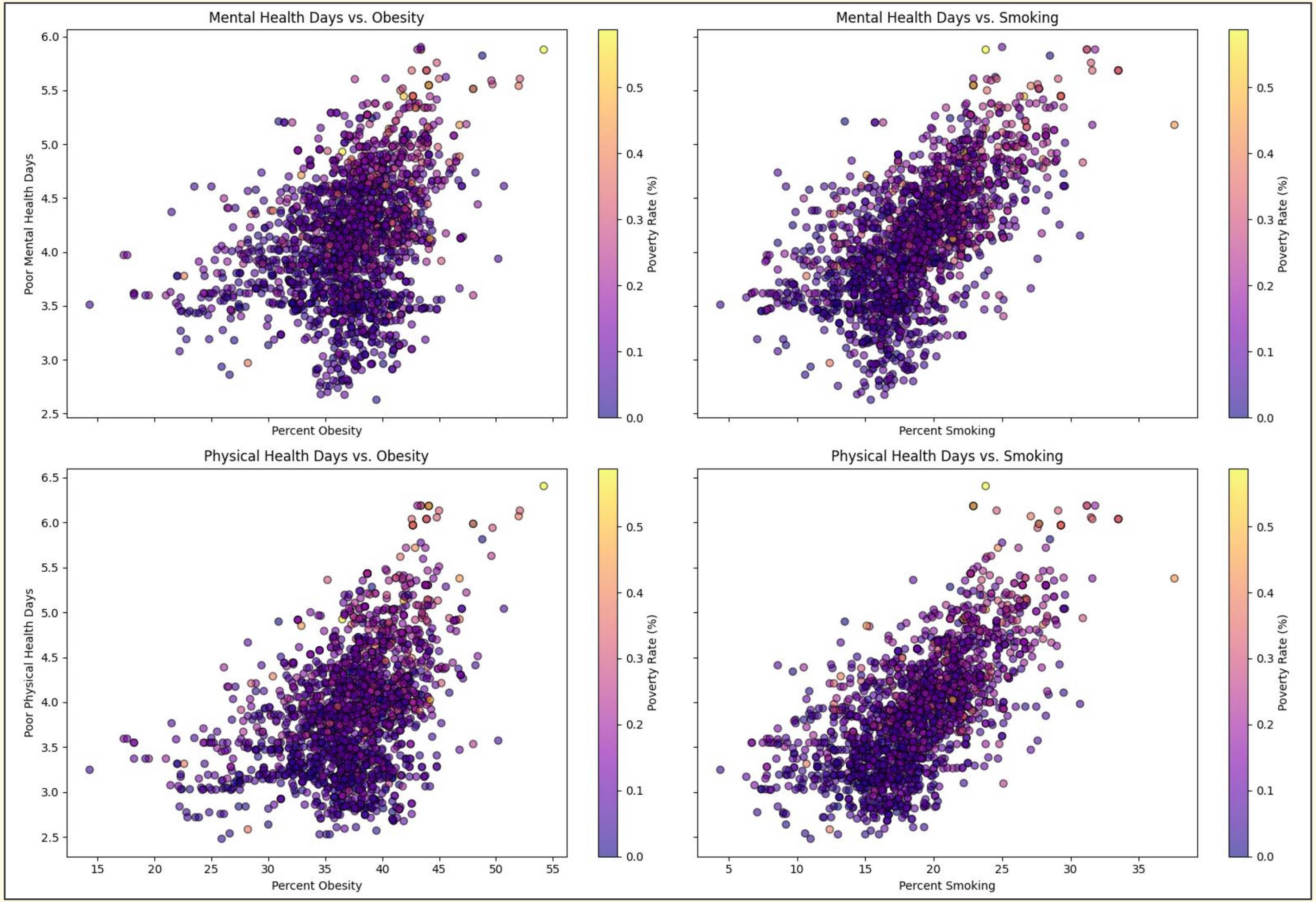
**

SMOKING & OBESITY

VS.

MENTAL & PHYSICAL HEALTH

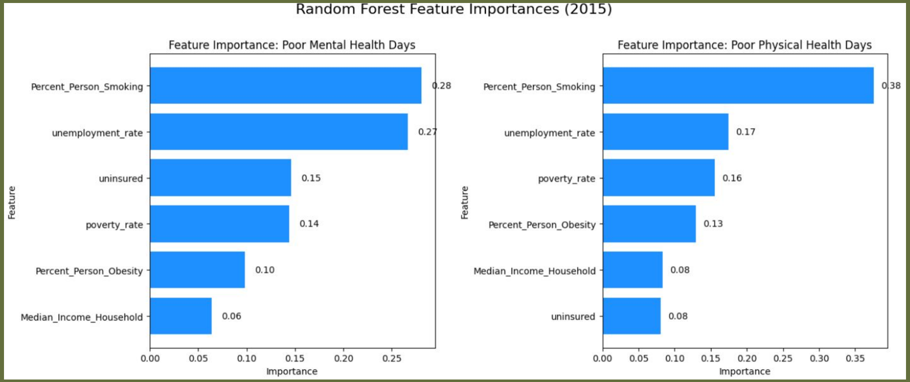
colored by **poverty rate**

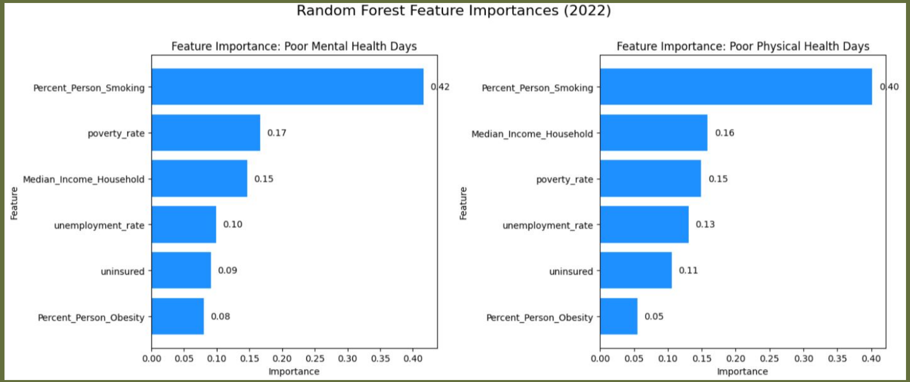


COUNTY-LEVEL ANALYSIS:

- **higher obesity** && **smoking rates** ⇒ more poor mental & physical health days
- **higher poverty (yellow)** ⇒ high risk of smoking, obesity && worse health outcomes

RANDOM FOREST FEATURE IMPORTANCES






```
Processing year: 2023
No data for year 2023
Processing year: 2024
No data for year 2024
Processing year: 2025
No data for year 2025
```

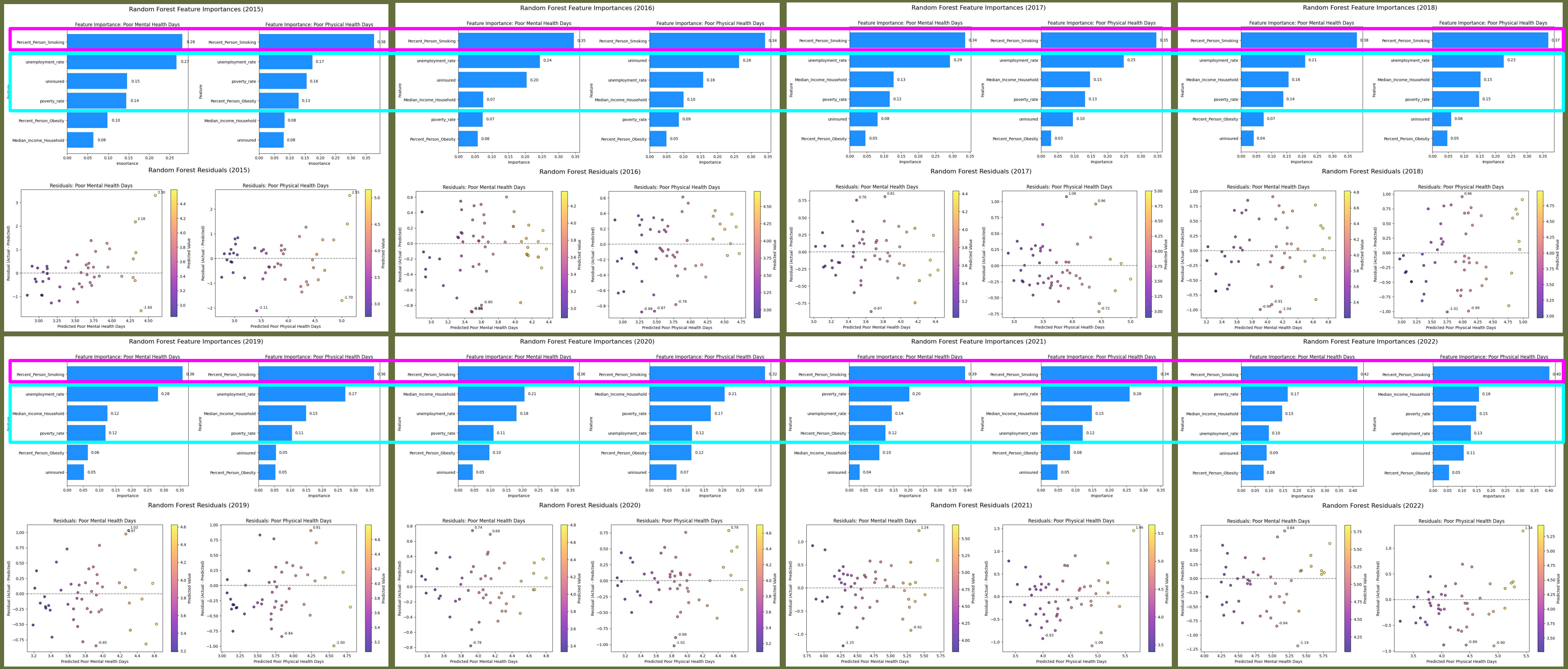
MENTAL & PHYSICAL HEALTH CONTRIBUTORS

Top Feature: Percent_Person_Smoking increases from 2015-2022.

Economic hardship (poverty, unemployment, low income) **consistently predicts worse health**.

Less important: lack of insurance & obesity rates are less predictive of poor health days.

RANDOM FOREST FEATURE IMPORTANCES && RESIDUALS SCATTERPLOT (2015-2022)




```
Processing year: 2023
No data for year 2023
Processing year: 2024
No data for year 2024
Processing year: 2025
No data for year 2025
```

Top Feature: Percent_Person_Smoking increases in importance from 2015-2022.

Economic hardship (poverty, unemployment, low income) **consistently predicts worse health**, but rankings for specific factors vary by year.

Less important is lack of insurance & obesity rates — less predictive of poor health days.

DISCUSSION &Next Steps

Smoking prevention and **rehabilitation** could have the biggest impact on reducing the frequency of poor mental & physical health days.

Economic improvements (addressing poverty, unemployment, and income) **remain important**, but their relative influence shifts with time.

Policy and intervention focus should **prioritize smoking reduction**, alongside **economic support**, for community health improvement.

01 IMPROVE DATA QUALITY

comprehensive data leads to more accurate and trustworthy analysis, pattern recognition, predictions

02 EXPLORE ADDITIONAL VARIABLES & FACTORS

might improve model performance or explain data variance and trends

03 EXAMINE CAUSALITY BETWEEN PATTERNS

understand reasons for observed associations, perceived relationships & trends