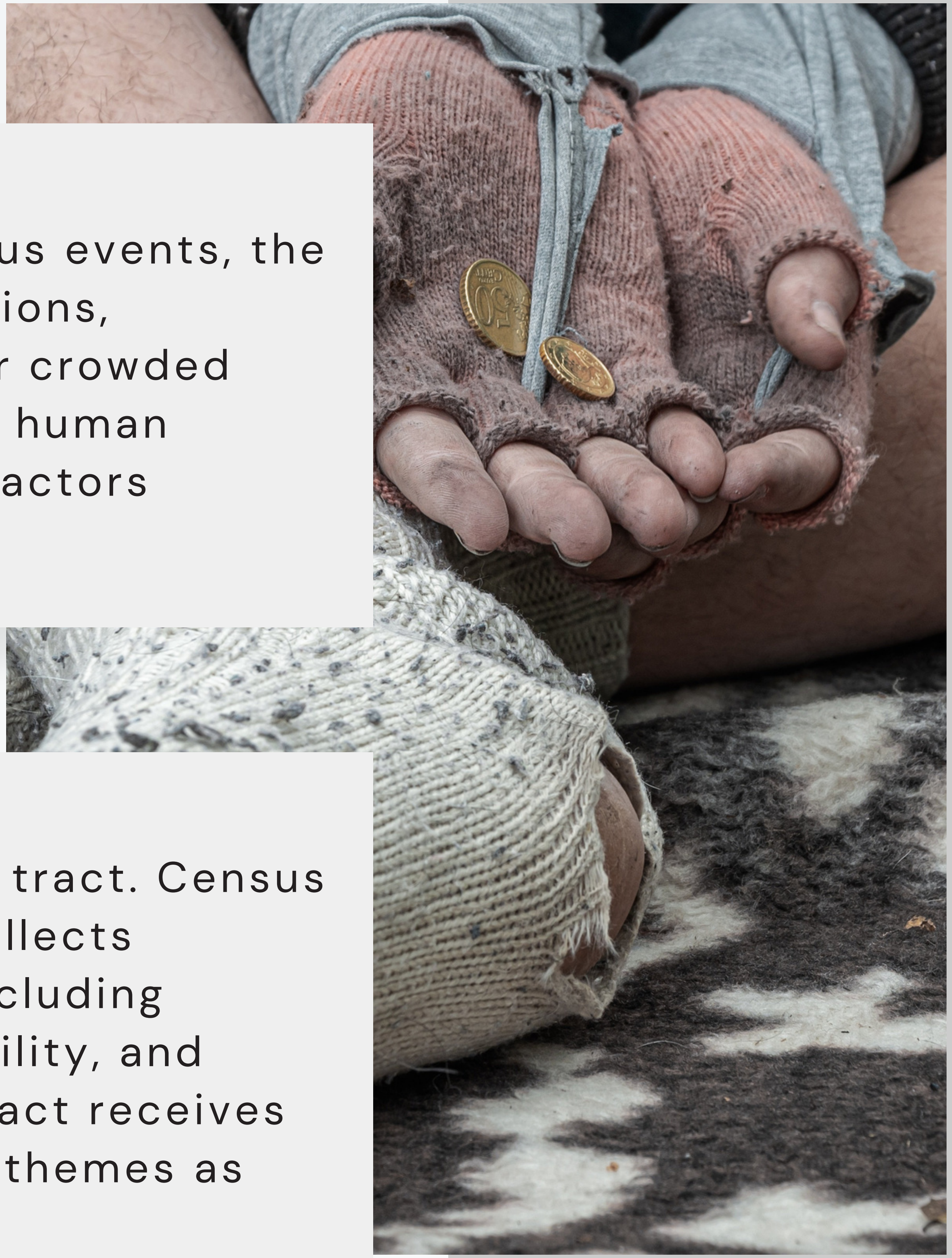
CDC/ATSDR
SOCIAL VULNERABILITY INDEX 2020

INTRODUCCIÓN



What is Social Vulnerability?
Every community must prepare for and respond to hazardous events, the degree to which a community exhibits certain social conditions, including high poverty, low percentage of vehicle access, or crowded households, may affect that community's ability to prevent human suffering and financial loss in the event of disaster. These factors describe a community's social vulnerability.

What is Social Vulnerability Index?
SVI indicates the relative vulnerability of every U.S. Census tract. Census tracts are subdivisions of counties for which the Census collects statistical data. SVI ranks the tracts on 16 social factors, including unemployment, racial and ethnic minority status, and disability, and further groups them into four related themes. Thus, each tract receives a ranking for each Census variable and for each of the four themes as well as an overall ranking.

OBJECTIVES

1. Identify Areas

Identify areas of high social vulnerability using the SVI to identify and visualize areas with a high index of social vulnerability. This can help local authorities and planners focus their efforts on preparation and response in these areas.

2. Model

Develop a predictive model using machine learning techniques that relate SVI to the probability of hazardous events, this can assist in the early identification of high-risk areas.

HYPOTHESIS

Null Hypothesis (H0)

There is no significant relationship between the independent variables (EP_NOHSDP, EP_UNEMP, EP_AGE65, EP_AGE17, EP_SNGPNT, EP_LIMENG, EP_MOBILE, EP_NOVEH) and the dependent variable (EP_POV150).

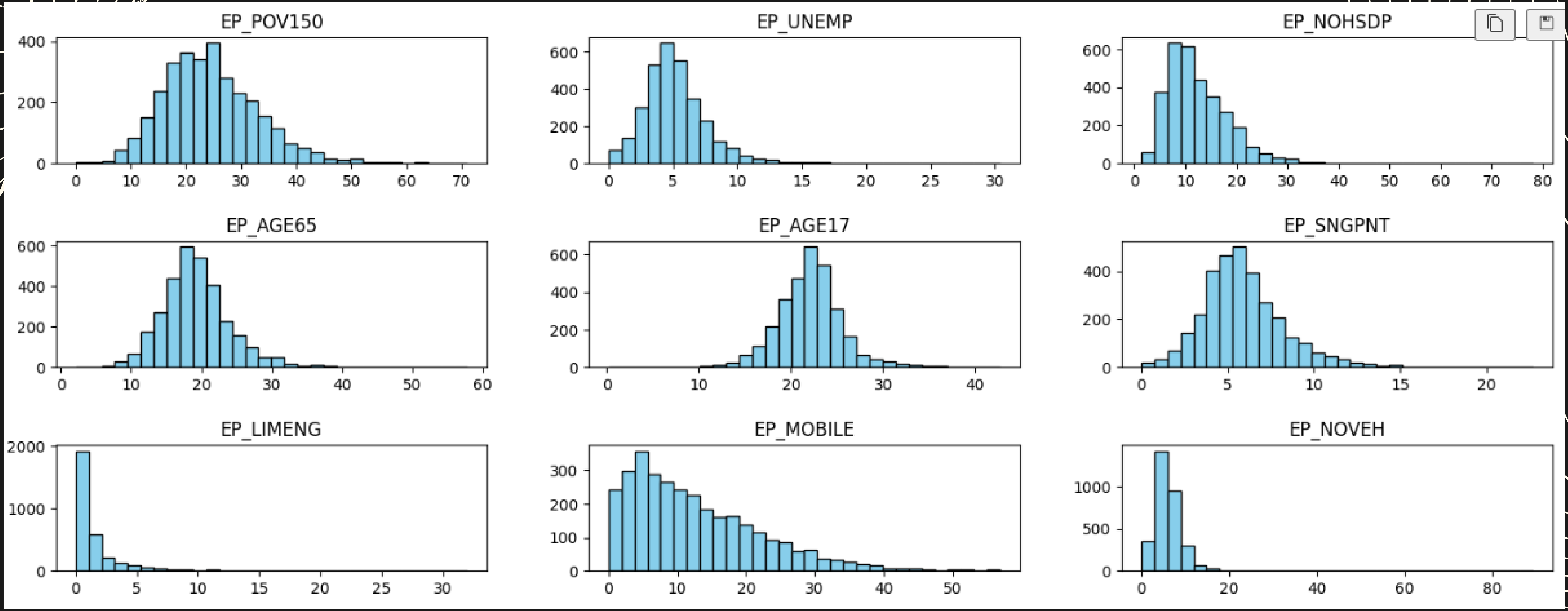
Alternative Hypothesis (H1)

There is a significant relationship between at least one of the independent variables and the dependent variable. In other words, at least one of the independent variables has a significant impact on the dependent variable.

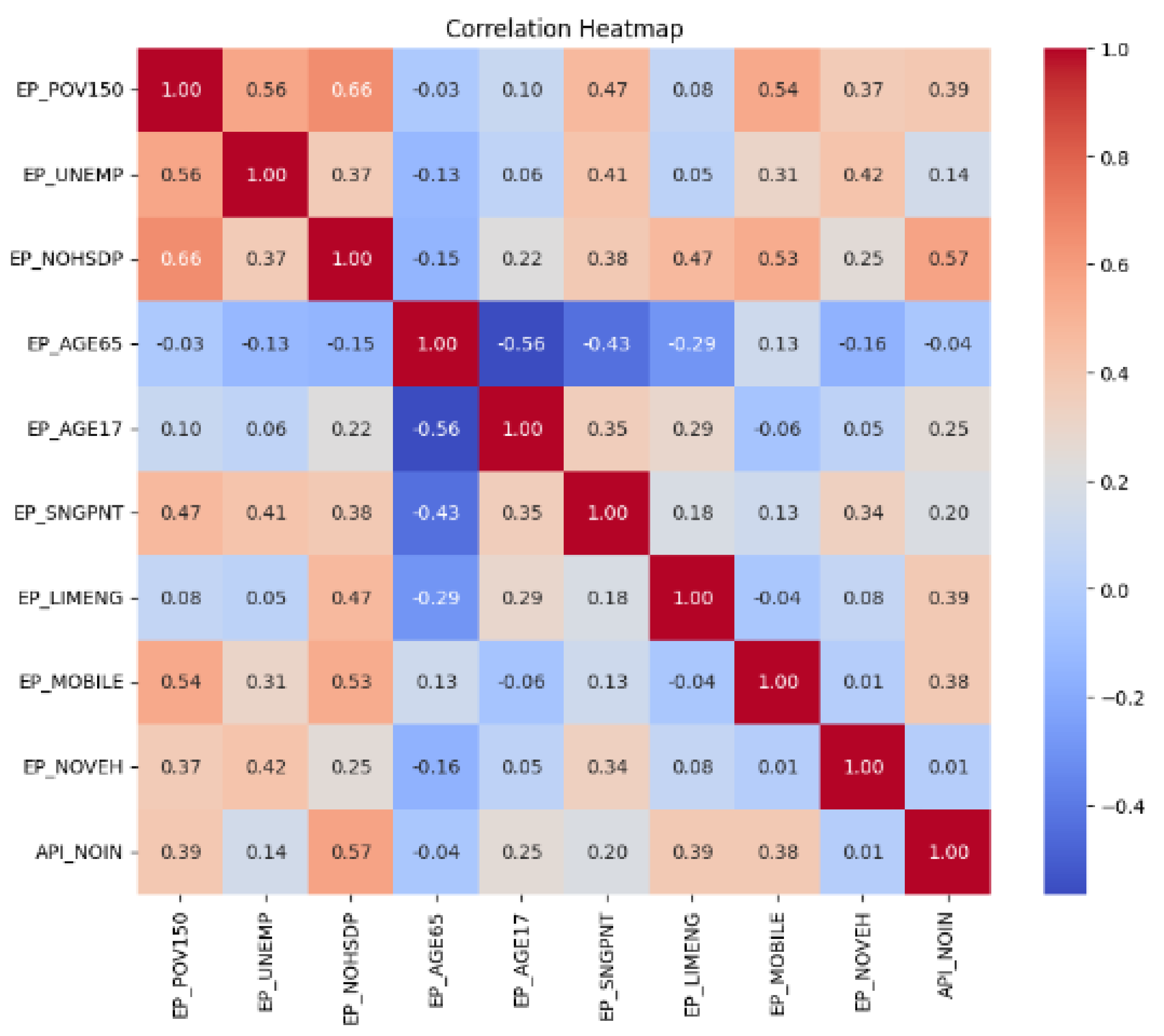
VARIABLES OF INTEREST

- STATE   State name
- EP_POV150 Percentage of persons below 150% poverty estimate
- EP_UNEMP Unemployment Rate estimate
- EP_NOHSDP Percentage of persons with no high school diploma (age 25+) estimate
- EP_AGE65 Percentage of persons aged 65 and older estimate, 2016-2020 ACS
- EP_AGE17 Percentage of persons aged 17 and younger estimate, 2016-2020 ACS
- EP_SNGPNT Percentage of single-parent households with children under 18 estimate, 2016-2020 ACS
- EP_LIMENG Percentage of persons (age 5+) who speak English "less than well" estimate, 2016-2020 ACS
- EP_MOBILE Percentage of mobile homes estimate
- EP_NOVEH Percentage of households with no vehicle available estimate
- API_NOIN Percentage of persons withot health insurance

DESCRIPTIVE CHARTS



DESCRIPTIVE CHARTS



METHODS

01

LINEAL REGRESION

02

K-NN

03

RANDOM FOREST

04

XGBOOST

CONCLUSION

First of all, the Null Hypothesis (HO) is rejected because all the independent variables have a significant relationship with the independent variable.

Speaking in terms of this study, it means that the Percentage of persons below 150% poverty estimate is strongly related to the rest of the variables.

In conclusion, and taking into account all the algorithms and methods used in this work, it can be concluded that the method that best fits the variables established in this data frame is XGBoost, with the best results in cross-validation of both MSE and R2.