Table 1. County Names Associated with unique Residence Code provided by NIJ

|  |  |  |
| --- | --- | --- |
| Unique  Residence Code (NIJ) | PUMA (combined) | Associated County Names |
| 1 | 1003, 4400 | Fulton, Douglas |
| 2 | 1008, 4300 | DeKalb,Newton, Rockdale |
| 3 | 1200, 1300 | Appling, Evans,Jeff, Davis, Montgomery, Tattnall, Telfair, Toombs,Wayne, Wheeler, Bleckley, Candler, Dodge, Emanuel, Johnson, Laurens, Treutlen, Wilcox |
| 4 | 1400, 1500, 1600 | Bibb, Houston, Pulaski, Baldwin, Crawford, Jones, Monroe, Peach,Putnam, Twiggs, Wilkinson |
| 5 | 1700, 1800 | Chattahoochee, Muscogee, Clay, Crisp, Dooly, Harris, Macon, Marion, Quitman, Randolph, Schley, Stewart |
| 6 | 2001, 2002, 2003, 4005 | DeKalb,Gwinnett |
| 7 | 100, 200, 500 | Camden, Glynn, McIntosh, Bryan, Liberty, Long,Atkinson, Bacon, Brantley, Charlton, Clinch |
| 8 | 4000, 4100, 4200 | Richmond, Columbia, Burke, Glascock, Hancock, Jefferson, Jenkins, Lincoln, McDuffle, Taliaferro, Warren, Washington |
| 9 | 5001, 6001, 6002 | Clayton |
| 10 | 2400, 5002 | Fayette,Clayton |
| 11 | 1001, 3004, 4600 | Fulton, Cobb |
| 12 | 1002, 1005, 3300, 3400, 4001, 4002, 4006 | Fulton, Forsyth,Hall, Gwinnett |
| 13 | 3101, 3102 | Cherokee |
| 14 | 1900, 3900, 4003, 4004 | Butts, Lamar, Pike, Spalding, Upson, Jasper, Morgan, Walton, Gwinnett |
| 15 | 3001, 3002, 3003, 3005 | Cobb |
| 16 | 2500, 4500 | Floyd, Haralson,Polk, Paulding |
| 17 | 2800, 2900, 3200, 3500 | Fannin, Gilmer, Gordon, Murray, Pickens, Bartow, Dawson, Lumpkin, Rabun, Towns, Union, White, Banks, Franklin, Habersham, Hart, Stephens |
| 18 | 600, 700, 800 | Lowndes, Ben Hill, Berrien, Brooks, Cook, Irwin, Tift, Turner, Colquitt, Thomas, Worth |
| 19 | 900, 1100 | Dougherty, Lee, Baker, Calhoun, Decatur, Early, Grady, Miller, Mitchell, Seminole, Terrell |
| 20 | 300, 401, 402 | Bulloch, Effingham, Screven, Chatham |
| 21 | 1004, 2100 | Fulton, Coweta |
| 22 | 2200, 2300 | Heard, Meriwether, Troup, Carroll |
| 23 | 1006, 1007, 2004 | Fulton, Dekalb |
| 24 | 2600, 2700 | Catoosa, Chattooga, Dade, Walker, Whitfield |
| 25 | 3600, 3700, 3800 | Clarke, Elbert, Greene, Madison, Oconee, Oglethorpe, Barrow, Jackson |

Table 2. Process information for all variables in the training and test datasets provided by NIJ

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Variable Name | Type | Number of Categories | Process Applied | | | | Number of Constructed Variables |
|  |  |  |  | Dummy Coding | One-hot encoding | Polynomial contrast coding | Numerical Assignment |  |
| 1 | Gender | Binary | 2 | x |  |  |  | 1 |
| 2 | Race | Binary | 2 | x |  |  |  | 1 |
| 3 | Age\_at\_Release | Ordinal | 7 |  | x | x | x | 14 |
| 4 | PUMAs | Nominal | 25 |  | x |  |  | 25 |
| 5 | Gang\_Affiliated | Binary | 2 | x |  |  |  | 1 |
| 6 | Supervision\_Risk\_Score\_First | Numeric |  |  |  |  | x | 1 |
| 7 | Supervision\_Level\_First | Ordinal | 3 |  | x | x |  | 5 |
| 8 | Education\_Level | Ordinal | 3 |  | x | x | x | 6 |
| 9 | Dependents | Ordinal | 4 |  | x |  | x | 5 |
| 10 | Prison\_Offence | Nominal | 5 |  | x |  |  | 5 |
| 11 | Prison\_Years | Ordinal | 4 |  | x | x |  | 7 |
| 12 | Prior\_Arrest\_Episodes\_Felony | Numeric |  | x |  |  | x | 2 |
| 13 | Prior\_Arrest\_Episodes\_Misd | Numeric |  | x |  |  | x | 2 |
| 14 | Prior\_Arrest\_Episodes\_Violent | Numeric |  | x |  |  | x | 2 |
| 15 | Prior\_Arrest\_Episodes\_Property | Numeric |  | x |  |  | x | 2 |
| 16 | Prior\_Arrest\_Episodes\_Drug | Numeric |  | x |  |  | x | 2 |
| 17 | Prior\_Arrest\_Episodes\_DVCharges | Numeric |  | x |  |  |  | 1 |
| 18 | Prior\_Arrest\_Episodes\_GunCharges | Numeric |  | x |  |  |  | 1 |
| 19 | Prior\_Conviction\_Episodes\_Felony | Numeric |  | x |  |  | x | 2 |
| 20 | Prior\_Conviction\_Episodes\_Misd | Numeric |  | x |  |  | x | 2 |
| 21 | Prior\_Conviction\_Episodes\_Violent | Numeric |  | x |  |  |  | 1 |
| 22 | Prior\_Conviction\_Episodes\_Property | Numeric |  | x |  |  | x | 2 |
| 23 | Prior\_Conviction\_Episodes\_Drug | Numeric |  | x |  |  | x | 2 |
| 24 | X\_v1 | Numeric |  | x |  |  | x | 2 |
| 25 | X\_v2 | Binary | 2 | x |  |  |  | 1 |
| 26 | X\_v3 | Binary | 2 | x |  |  |  | 1 |
| 27 | X\_v4 | Binary | 2 | x |  |  |  | 1 |
| 28 | Prior\_Revocations\_Parole | Binary | 2 | x |  |  |  | 1 |
| 29 | Prior\_Revocations\_Probation | Binary | 2 | x |  |  |  | 1 |
| 30 | Condition\_MH\_SA | Binary | 2 | x |  |  |  | 1 |
| 31 | Condition\_Cog\_Ed | Binary | 2 | x |  |  |  | 1 |
| 32 | Condition\_Other | Binary | 2 | x |  |  |  | 1 |
| 33 | Violations\_ElectronicMonitorin | Binary | 2 | x |  |  |  | 1 |
| 34 | Violations\_InstructionsNotFollowed | Binary | 2 | x |  |  |  | 1 |
| 35 | Violations\_FailtoReport | Binary | 2 | x |  |  |  | 1 |
| 36 | Violations\_MoveWithoutPermission | Binary | 2 | x |  |  |  | 1 |
| 37 | Delinquency\_Reports | Numeric |  | x |  |  | x | 2 |
| 38 | Program\_Attendances | Numeric |  | x |  |  | x | 2 |
| 39 | Program\_UnexcusedAbsences | Numeric |  | x |  |  | x | 2 |
| 40 | Residence\_Changes | Numeric |  | x |  |  | x | 2 |
| 41 | Avg\_Days\_per\_DrugTest | Numeric |  |  |  |  | x | 1 |
| 42 | DrugTests\_THC\_Positive | Numeric |  |  |  |  | x | 1 |
| 43 | DrugTests\_Cocaine\_Positive | Numeric |  |  |  |  | x | 1 |
| 44 | DrugTests\_Meth\_Positive | Numeric |  |  |  |  | x | 1 |
| 45 | DrugTests\_Other\_Positive | Numeric |  |  |  |  | x | 1 |
| 46 | Percent\_Days\_Employed | Numeric |  |  |  |  | x | 1 |
| 47 | Jobs\_Per\_Year | Numeric |  |  |  |  | x | 1 |
| 48 | Employment\_Exempt | Binary | 2 | x |  |  |  | 1 |

*Notes.* The variables are listed in order they appear in the training dataset provided by NIJ. A total of 48 predictors are recoded into a total of 122 variables after processing all variables. In addition to these 122 variables, a Principal Component Analysis was run for crime related variables. PCA revealed that these variables can be grouped into four categories. Therefore, an additional four composite variables were created as basic sum score of the crime related variables in these four categories. The R code that is used to process these variables for more detailed information can be found at this link (<https://github.com/czopluoglu/nij-competition/blob/main/R/03_data%20prep.r>).

Table 3. List of variables aggregated at the county-level from 2018 American Community Survey (ACS) 5-year estimates

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Variable Name | Type | Number of Categories | Process Applied | | | | Number of Constructed Variables |
|  |  |  |  | Dummy Coding | One-hot encoding | Polynomial contrast coding | Numerical Assignment |  |
| 1 | ACCESS | Nominal | 3 | x |  |  |  | 1 |
| 2 | ACR | Nominal | 3 |  | x |  |  | 3 |
| 3 | AGEP | Numeric |  |  |  |  | x | 1 |
| 4 | AGS | Ordinal | 6 |  | x |  |  | 6 |
| 5 | BATH | Binary | 2 | x |  |  |  | 1 |
| 6 | BLD | Nominal | 10 |  | x |  |  | 10 |
| 7 | BROADBND | Binary | 2 | x |  |  |  | 1 |
| 8 | BUS | Binary | 2 | x |  |  |  | 1 |
| 9 | CIT | Nominal | 5 |  | x |  |  | 5 |
| 10 | COMPOTHX | Binary | 2 | x |  |  |  | 1 |
| 11 | CONP (log transformed) | Numeric |  |  |  |  | x | 1 |
| 12 | COW | Nominal | 9 |  | x |  |  | 9 |
| 13 | DDRS | Binary | 2 | x |  |  |  | 1 |
| 14 | DEAR | Binary | 2 | x |  |  |  | 1 |
| 15 | DEYE | Binary | 2 | x |  |  |  | 1 |
| 16 | DIALUP | Binary | 2 | x |  |  |  | 1 |
| 17 | DIS | Binary | 2 | x |  |  |  | 1 |
| 18 | DOUT | Binary | 2 | x |  |  |  | 1 |
| 19 | DPHY | Binary | 2 | x |  |  |  | 1 |
| 20 | DRAT | Ordinal |  |  |  |  | x | 2 |
| 21 | DRATX | Binary | 2 | x |  |  |  | 1 |
| 22 | DREM | Binary | 2 | x |  |  |  | 1 |
| 23 | ELEFP | Nominal | 3 |  | x |  |  | 3 |
| 24 | ELEP | Numeric |  |  |  |  | x | 1 |
| 25 | ENG | Ordinal |  |  |  |  | x | 1 |
| 26 | FER | Binary | 2 | x |  |  |  | 1 |
| 27 | FES | Nominal | 8 |  | x |  |  | 8 |
| 28 | FINCP | Numeric |  |  |  |  | x | 1 |
| 29 | FPARC | Nominal | 4 |  | x |  |  | 4 |
| 30 | FS | Binary | 2 | x |  |  |  | 1 |
| 31 | FULP (log transformed) | Numeric |  |  |  |  | x | 1 |
| 32 | GASP (log transformed) | Numeric |  |  |  |  | x | 1 |
| 33 | GCL | Binary | 2 | x |  |  |  | 1 |
| 34 | GCM | Ordinal |  | x |  |  |  | 5 |
| 35 | GCR | Binary | 2 | x |  |  |  | 1 |
| 36 | GRNTP | Numeric |  |  |  |  | x | 2 |
| 37 | GRPIP | Numeric |  |  |  |  | x | 1 |
| 38 | HFL | Nominal | 9 |  | x |  |  | 9 |
| 39 | HHL | Nominal | 5 |  | x |  |  | 5 |
| 40 | HHT | Nominal | 7 |  | x |  |  | 7 |
| 41 | HINCP | Numeric |  |  |  |  | x | 1 |
| 42 | HINS1 | Binary | 2 | x |  |  |  | 1 |
| 43 | HINS2 | Binary | 2 | x |  |  |  | 1 |
| 44 | HINS3 | Binary | 2 | x |  |  |  | 1 |
| 45 | HINS4 | Binary | 2 | x |  |  |  | 1 |
| 46 | HINS5 | Binary | 2 | x |  |  |  | 1 |
| 47 | HINS6 | Binary | 2 | x |  |  |  | 1 |
| 48 | HINS7 | Binary | 2 | x |  |  |  | 1 |
| 49 | HISPEED | Binary | 2 | x |  |  |  | 1 |
| 50 | HUGCL | Binary | 2 | x |  |  |  | 1 |
| 51 | HUPAC | Nominal | 4 |  | x |  |  | 4 |
| 52 | HUPAOC | Nominal | 4 |  | x |  |  | 4 |
| 53 | HUPARC | Nominal | 4 |  | x |  |  | 4 |
| 54 | INSP | Numeric |  | x |  |  |  | 1 |
| 55 | INTP | Numeric |  |  |  |  | x | 1 |
| 56 | JWMNP | Numeric |  |  |  |  | x | 1 |
| 57 | JWRIP | Numeric |  |  |  |  | x | 1 |
| 58 | JWTR | Nominal | 12 | x |  |  |  | 6 |
| 59 | KIT | Binary | 2 | x |  |  |  | 1 |
| 60 | LANX | Binary | 2 | x |  |  |  | 1 |
| 61 | LAPTOP | Binary | 2 | x |  |  |  | 1 |
| 62 | LNGI | Binary | 2 | x |  |  |  | 1 |
| 63 | MAR | Nominal | 5 |  | x |  |  | 5 |
| 64 | MARHD | Binary | 2 | x |  |  |  | 1 |
| 65 | MARHM | Binary | 2 | x |  |  |  | 1 |
| 66 | MARHT | Ordinal | 3 |  | x |  |  | 3 |
| 67 | MARHW | Binary | 2 | x |  |  |  | 1 |
| 68 | MHP | Numeric |  |  |  |  | x | 1 |
| 69 | MIG | Binary | 2 | x |  |  |  | 1 |
| 70 | MIL | Nominal | 4 |  | x |  |  | 4 |
| 71 | MLPA | Binary | 2 | x |  |  |  | 1 |
| 72 | MLPB | Binary | 2 | x |  |  |  | 1 |
| 73 | MLPCD | Binary | 2 | x |  |  |  | 1 |
| 74 | MLPE | Binary | 2 | x |  |  |  | 1 |
| 75 | MLPFG | Binary | 2 | x |  |  |  | 1 |
| 76 | MLPH | Binary | 2 | x |  |  |  | 1 |
| 77 | MLPI | Binary | 2 | x |  |  |  | 1 |
| 78 | MLPJ | Binary | 2 | x |  |  |  | 1 |
| 79 | MLPK | Binary | 2 | x |  |  |  | 1 |
| 80 | MRGI | Binary | 2 | x |  |  |  | 1 |
| 81 | MRGP | Numeric |  |  |  |  | x | 1 |
| 82 | MRGT | Binary | 2 | x |  |  |  | 1 |
| 83 | MRGX | Nominal | 3 |  | x |  |  | 3 |
| 84 | MSP | Nominal | 6 |  | x |  |  | 6 |
| 85 | MULTG | Binary | 2 | x |  |  |  | 1 |
| 86 | MV | Nominal | 7 |  | x |  |  | 7 |
| 87 | NATIVITY | Binary | 2 | x |  |  |  | 1 |
| 88 | NOC | Numeric |  |  |  |  | x | 1 |
| 89 | NP | Numeric |  |  |  |  | x | 1 |
| 90 | NPF | Numeric |  |  |  |  | x | 1 |
| 91 | NPP | Numeric |  |  |  |  | x | 1 |
| 92 | NR | Binary | 2 | x |  |  |  | 1 |
| 93 | NRC | Numeric |  |  |  |  | x | 1 |
| 94 | NWAB | Nominal | 3 |  | x |  |  | 3 |
| 95 | NWAV | Nominal | 5 |  | x |  |  | 5 |
| 96 | NWLA | Nominal | 3 |  | x |  |  | 3 |
| 97 | NWLK | Nominal | 3 |  | x |  |  | 3 |
| 98 | NWRE | Nominal | 4 |  | x |  |  | 3 |
| 99 | OC | Binary | 2 | x |  |  |  | 1 |
| 100 | OCPIP | Numeric |  |  |  |  | x | 2 |
| 101 | OIP | Numeric |  |  |  |  | x | 1 |
| 102 | OTHSVCEX | Binary | 2 | x |  |  |  | 1 |
| 103 | PAOC | Nominal | 4 |  | x |  |  | 4 |
| 104 | PAP | Numeric |  |  |  |  | x | 1 |
| 105 | PARTNER | Nominal | 5 |  | x |  |  | 5 |
| 106 | PERNP | Numeric |  |  |  |  | x | 1 |
| 107 | PINCP | Numeric |  |  |  |  | x | 1 |
| 108 | PLM | Binary | 2 | x |  |  |  | 1 |
| 109 | POVPIP | Numeric |  |  |  |  | x | 1 |
| 110 | PRIVCOV | Binary | 2 | x |  |  |  | 1 |
| 111 | PSF | Binary | 2 | x |  |  |  | 1 |
| 112 | PUBCOV | Binary | 2 | x |  |  |  | 1 |
| 113 | R18 | Binary | 2 | x |  |  |  | 1 |
| 114 | R60 | Ordinal | 3 |  | x |  |  | 3 |
| 115 | R65 | Ordinal | 3 |  | x |  |  | 3 |
| 116 | RACAIAN | Binary | 2 | x |  |  |  | 1 |
| 117 | RACASN | Binary | 2 | x |  |  |  | 1 |
| 118 | RACBLK | Binary | 2 | x |  |  |  | 1 |
| 119 | RACWHT | Binary | 2 | x |  |  |  | 1 |
| 120 | RC | Binary | 2 | x |  |  |  | 1 |
| 121 | REFR | Binary | 2 | x |  |  |  | 1 |
| 122 | RETP | Numeric |  |  |  |  | x | 1 |
| 123 | RMSP | Numeric |  |  |  |  | x | 1 |
| 124 | RNTM | Binary | 2 | x |  |  |  | 1 |
| 125 | RWAT | Binary | 2 | x |  |  |  | 1 |
| 126 | SATELLITE | Binary | 2 | x |  |  |  | 1 |
| 127 | SCIENGP | Binary | 2 | x |  |  |  | 1 |
| 128 | SCIENGRLP | Binary | 2 | x |  |  |  | 1 |
| 129 | SEMP | Numeric |  |  |  |  | x | 1 |
| 130 | SINK | Binary | 2 | x |  |  |  | 1 |
| 131 | SMARTPHONE | Binary | 2 | x |  |  |  | 1 |
| 132 | SMOCP | Numeric |  |  |  |  | x | 1 |
| 133 | SMP | Numeric |  |  |  |  | x | 1 |
| 134 | SMX | Nominal | 4 |  | x |  |  | 4 |
| 135 | SRNT | Binary | 2 | x |  |  |  | 1 |
| 136 | SSIP | Numeric |  |  |  |  | x | 1 |
| 137 | SSMC | Ordinal | 3 | x |  |  |  | 1 |
| 138 | SSP | Numeric |  |  |  |  | x | 1 |
| 139 | STOV | Binary | 2 | x |  |  |  | 1 |
| 140 | SVAL | Binary | 2 | x |  |  |  | 1 |
| 141 | TABLET | Binary | 2 | x |  |  |  | 1 |
| 142 | TAXAMT | Numeric |  |  |  |  | x | 1 |
| 143 | TEL | Binary | 2 | x |  |  |  | 1 |
| 144 | TEN | Nominal | 4 |  | x |  |  | 4 |
| 145 | TOIL | Binary | 2 | x |  |  |  | 1 |
| 146 | VALP | Numeric |  |  |  |  | x | 1 |
| 147 | VEH | Numeric |  |  |  |  | x | 1 |
| 148 | WAGP | Numeric |  |  |  |  | x | 1 |
| 149 | WATFP | Nominal | 3 |  | x |  |  | 3 |
| 150 | WATP | Numeric |  |  |  |  | x | 1 |
| 151 | WGTP | Numeric |  |  |  |  | x | 1 |
| 152 | WIF | Ordinal | 4 |  | x |  |  | 4 |
| 153 | WKHP | Numeric |  |  |  |  | x | 1 |
| 154 | WKL | Nominal | 3 |  | x |  |  | 3 |
| 155 | WKW | Ordinal |  |  |  |  | x | 1 |
| 156 | WRK | Binary | 2 | x |  |  |  | 1 |
| 157 | YBL | Ordinal | 22 |  |  |  | x | 1 |

*Notes.* The detailed codebook about these variables can be found in the 2014 – 2018 ACS PUMA Data Dictionary (<https://www2.census.gov/programs-surveys/acs/tech_docs/pums/data_dict/PUMS_Data_Dictionary_2014-2018.pdf>). A total of 157 predictors are recoded into a total of 295 variables after processing all variables. In addition, a Principal Component Analysis was run for all 295 variables. Standardized composite scores for the first four principal component were added to the dataset. The R code that is used to download the 2018 ACS database and process these variables can be found at this link <https://github.com/czopluoglu/nij-competition/blob/main/R/02_geodata.r> ).